The Dissertation Committee for Joseph T. Abatemarco Certifies that this is the approved version of the following dissertation:

Novel Approaches for the Evolutionary Engineering of Pathways in *Saccharomyces Cerevisiae*

Committee:

Hal Alper, Supervisor

Jeffrey Barrick

Lydia Contreras

Andrew Ellington

Jennifer Maynard

Novel Approaches for the Evolutionary Engineering of Pathways in *Saccharomyces Cerevisiae*

by

Joseph T. Abatemarco, BS Chem & Bio Engr

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2016

Dedication

To my family

Acknowledgements

I would first like to acknowledge and thank my advisor, Hal Alper. He has been an incredible source of ideas, advice, and inspiration throughout my time in graduate school. The support, communication and feedback he gives his students set him apart from many other graduate advisors and managers, and I could not be happier that I joined his lab when I began at UT.

I would also like to thank my committee members, Jeffrey Barrick, Lydia Contreras, Andrew Ellington, and Jennifer Maynard. I have had interesting and useful conversations with each of them over the course of my time at UT, and they have all been a huge help in offering advice and suggestions, both before and after my initial thesis proposal. These ideas helped me refine and carry out my research goals, and I would not have finished without their support.

Next I need to acknowledge the other members of Alper lab I had the pleasure of working with during my graduate studies. This lab has been a wonderful group to be a part of, especially due to its friendly and collaborative atmosphere. Impromptu discussions in the lab or in the office led to many great insights and ideas that had real impact on my work. Lab members have also been enormously helpful during group presentations, interrupting often with clarifying questions, suggestions, and ideas. This type of feedback was essential to growth as a researcher, and there could not have been a better group to give it. Specifically, I'd like to thank a few members who directly contributed to my work here. Nathan Crook and Jie Sun were my two collaborators on the ICE project, and together we formed an excellent team. Through long discussions, help with experiments, back-and-forth editing, and travel to several conferences, we were

able to accomplish some excellent new research. I'd next like to thank John Leavitt, who helped me enormously with the work on aromatic amino acids. He explained and summarized the state of the field clearly, and donated many strains and plasmids that ended up being necessary for this research. He also developed the tyrosine quantification assay on which I relied heavily for many applications. Next I need to acknowledge James Wagner, who worked with me on the ICE, RAPID, and 2A projects. His extremely knowledgeable advice and conversations were a huge help in refining these projects, and his contributions to each were extremely useful.

Outside of our group, I'd like to thank Dr. Howard Salis, who collaborated with us on RNA aptamer development, leading to huge benefits in turning these molecules into useful and applicable sensors; we could not have done this without his input. I'd also like to thank Dr. Adam Abate, with whom we worked to learn and apply microdroplet technology without which RAPID would not be possible. In addition, three post-docs from Dr. Abate's lab, Wafa Hassouneh, Maen Sarhan, and Leqian Liu, were all great partners in this work, and I learned a lot from working with each of them.

Next, I had the pleasure of working with several undergraduates that I would like to acknowledge. Alex Schmitz worked with me both on ICE and significant parts of the 2A project, and as an extremely competent and independent researcher. I am confident he will do very well in his own ongoing graduate studies. Richard Paul was another talented undergraduate who helped enormously with the aptamer development work, and who quickly learned about our research and how to apply these ideas. Su Min Cho worked for years on both the ICE and RAPID projects, making significant contributions to each. Finally, Julio Mejia was a very talented researcher who not only learned lab techniques quickly, but learned to think critically about the work we were doing. All four of these

students together taught me how to become a better teacher and leader, while at the same time making significant contributions to our research work.

Next I need to acknowledge some others who have made a personal impact during my time at UT. First, my wonderful teammates on our Austin Sports and Social dodgeball team, especially Bart Dear, who all helped make every Wednesday night game a great time, win or lose. I'd also to thank many of the other good friends I made in grad school, who made the time here something special and who I hope will continue to remain my friends as we all graduate and go our separate ways: Bart Dear, Ben Wendel, Sonali Chopra, Meghali Chopra, Amanda Paine, Michelle Dose, Bailee Roach, Austin Lane, Doug Pernik, Erik Johnson, Jovan Kamcev, Abby Ondeck, Ellen Wagner, Sean DeRosa, Reika Katsumata, Shuqi Zhang, Kelly Markham, Nick Morse, Emily Adkins, Matt Deaner, and James and Angela Wagner.

Lastly I'd like to acknowledge my family, who gave me unconditional support as I made the (to some, questionable) decision to leave a good job and move halfway across the country for graduate school. I need to especially thank my mom, who moved to Texas for an entire year to help when our son was born; without her I don't know how we would have managed! Our son JT himself, who is a constant joy and brightens every single day. Finally, I need to thank my wonderful wife Sharon – who came on this adventure with me, who supported me through it all, and who will always remain my partner and best friend. I love you.

Novel Approaches for the Evolutionary Engineering of Pathways in *Saccharomyces Cerevisiae*

Joseph T. Abatemarco, PhD

The University of Texas at Austin, 2016

Supervisor: Hal Alper

Modern biotechnological tools are making microbial production of chemicals, fuels, and pharmaceuticals increasingly practical and economically feasible. The field of metabolic engineering aims to enable this production by hijacking cellular systems to modify metabolism, converting each cell into an efficient chemical reactor. Traditionally, this has been accomplished through combining various knockouts and/or overexpressions of metabolic genes, but directed evolution strategies are often critical for improving metabolic pathways beyond native activity. Due to the complexity of cellular metabolism, simply evolving single genetic parts in a stepwise fashion can be limiting. In this work, we develop novel and powerful methods for applying directed evolution to the engineering of metabolic pathways in the yeast *Saccharomyces cerevisiae*.

First, we develop a method for *in vivo* Continuous Evolution (ICE), which uses a synthetic retrotransposon element to allow generation of the largest mutant libraries of any *in vivo* mutational generation approach in yeast. This method is then validated by using it to rapidly evolve a variety of diverse genetic systems, including single enzymes, global transcriptional regulators, and multi-gene pathways. Next, we apply a modeling approach to create novel biosensors that can rapidly screen for production and secretion phenotypes in these large mutant libraries. These biosensors are then imported into

microfluidic droplet systems to apply this screen in a high-throughput manner, creating a novel platform for screening libraries finally commensurate with the high level of diversity generation previously enabled. This method is validated by evolving an overproduction and secretion phenotype, resulting in strains that produce significantly elevated levels of aromatic amino acids. Finally, we develop and characterize new tools to enable the expression and evolution of multiple genes in a single genetic cassette. Taken together, these novel technologies significantly advance the state of the art for evolutionary engineering of metabolic pathways and will enable the evolution of pathways of enzymes for rapidly improving production of a number of desirable high-value biochemicals.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction [1]

Sustainable bio-based production of chemicals, fuels, and pharmaceuticals can be accomplished in part due to the extraordinary features and versatility of cellular metabolism. The field of metabolic engineering seeks to introduce genetic manipulations to rewire and reprogram microbes into cellular factories capable of producing a variety of high-value chemicals [1-5]. In doing so, metabolites must be rerouted, regulation must be altered, and new enzymatic pathways must be introduced and balanced. Traditionally, this rewiring has been accomplished through combining various knockouts and/or overexpressions of metabolic genes, resulting in an improved flux through either homologous or heterologous pathways. This approach has been highly successful in diversifying the chemical palate of a cell to produce key pharmaceuticals and nutraceuticals, fuels, commodity/specialty chemicals, polymer precursors and even complex, full-length polymers [6]. Although most metabolic engineering studies have utilized either *Escherichia coli* or *Saccharomyces cerevisiae* [7], there has been an increasing interest in exploring non-conventional hosts including fungi such as *Yarrowia lipolytica*, *Ashbya gossypii*, *Phaffia rhodozyma*, and *Hansenula polymorpha* [8]. Regardless of the host, the field has seen a rapid explosion in the number of systems-level approaches and molecular capabilities. Both synthetic and systems biology have improved our ability to tune cellular behavior aided by metabolic models, "omics" studies, and engineered synthetic parts such as promoters and transcription factor binding sites [9]. Additionally, the application of systems biology combined with metabolic engineering has enabled a Systems Metabolic Engineering paradigm that uses model-

[1] Adapted from Abatemarco, J., Hill, A., & Alper, H. S. (2013). *Expanding the metabolic engineering toolbox with directed evolution*. Biotechnology Journal, 8(12), 1397-1410. doi: 10.1002/biot.201300021. Joseph Abatemarco and Andrew Hill contributed equally to this work.

guided approaches and ‘omics data to rewire cellular systems [10, 11]. Regardless of the exact approach, diverse techniques improve the capability of metabolic engineering and enable microorganisms to sustainably produce high value chemicals from renewable substrates. However, there are occasions when metabolic engineering strategies require more than simple gene deletions and overexpressions. In this context, directed evolution is rapidly being adopted for many metabolic engineering and synthetic biology applications to introduce improved or novel functions that address complex cellular challenges.

1.1 Directed Evolution is Critical to Metabolic Engineering

Directed protein evolution has been applied in a variety of situations to effectively change the function of a starting protein or DNA scaffold. The basic premise behind this approach is that due to the complexity of protein dynamics, small changes in amino acid sequence can lead to radically different folding behavior, function, and/or net catalytic activity. Hence, the overall “evolvability” of any genetically encoded part can be exploited by creating a large library of mutants followed by screening for a variant possessing the desired function. While many approaches for diversification and selection have been described, the most common, unguided approach involves in vitro mutant generation using error-prone PCR followed by *in vivo* screening using a binding assay, fluorescence sorting screen, or serially culturing the cells in specific growth conditions [12].

However, a directed evolution approach is predicated upon two successful attributes: sufficient library diversity and an accurate, high-throughput screen. Library diversity is critical considering the vast, unsearchable number of possible mutations; in a

small polypeptide of only 100 amino acids, there are 20^{100} possible variants ($>10^{130}$). More sophisticated means of protein library design can be incorporated to enrich the library with more functional, desirable mutants, such as using models and/or phylogenetic relationships to identify specific regions or residues to target for mutagenesis [13-15]. In addition, iterative searches of protein-wide and focused mutagenesis can help identify key beneficial mutations that influence and optimize activity [16-18]. Taken together, these approaches can improve the quality of random libraries and increase the probability that a desirable mutation can be identified.

Beyond library generation, screen design is the current major limitation of directed evolution in the context of metabolic engineering. Although larger library sizes improve the chance of discovering beneficial mutations, current high-throughput screening techniques are limited in how meaningfully they characterize those mutants [19]. While growth-based screens are effective at improving catabolic enzymes, many metabolic enzymes of interest are not related to growth (and often, the growth objective and biotechnological objective are at odds). As a result, lower throughput assays such as GC, MS, or capillary electrophoresis are necessary to accurately determine a mutagenized enzyme's fitness [20], and thus place a severe limitation on the library size sampled using a directed evolution approach. The development of a generic, high-throughput and accurate assessment of desired protein functions is still lacking.

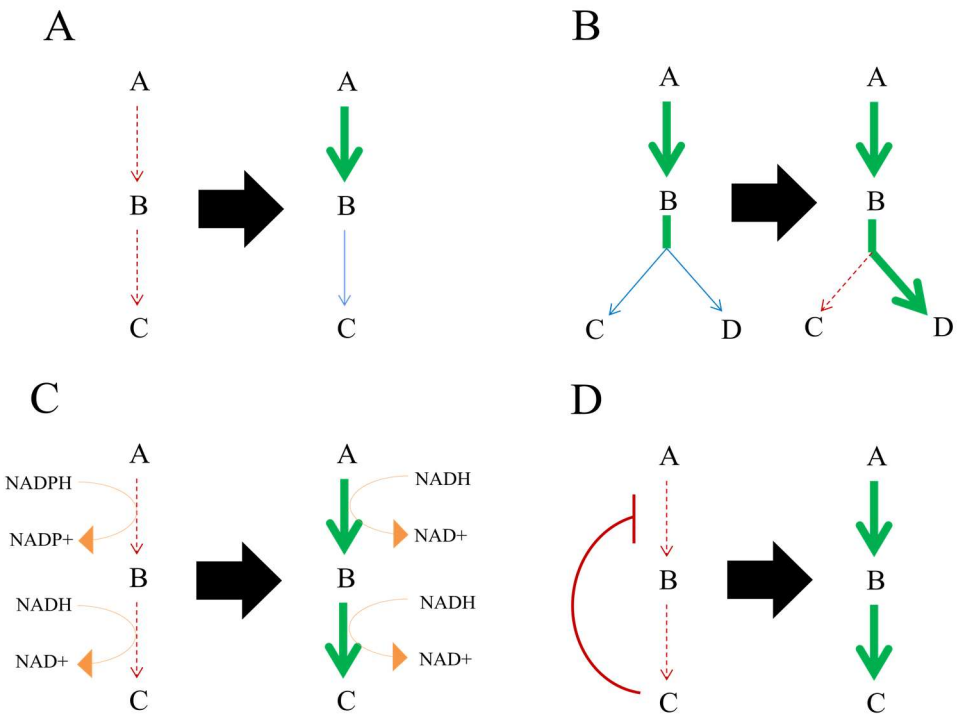
Despite these limitations, directed evolution has been applied in a number of metabolic engineering applications including identifying mutants that modify enzyme productivity [17, 21], substrate specificity [16, 22, 23], cofactor requirements [24] [25], and protein stability [26]. Furthermore, directed evolution has been applied to non-catalytic entities in an effort to modify metabolic regulation [27, 28], adjust binding kinetics [16] [29], and rapidly evolve the entire genome [30-34]. These examples are

among a growing list which highlights recent approaches of utilizing directed protein evolution to accomplish metabolic engineering goals.

1.1.1 Directed Evolution of Enzymes for Metabolic Engineering

As described above, one goal of metabolic engineering is to improve pathway flux which is usually accomplished through the amplification of enzyme levels. However, simple examination of Michaelis-Menten kinetics reveals at least two possible routes to improve overall productivity: (1) altering the amount of available enzyme or (2) altering the catalytic properties (ie k_{cat}/K_m). While the first alteration can be achieved through gene overexpression, the second type of alteration requires directed evolution and protein engineering to augment catalytic properties. To examine this further, consider a pathway exclusively controlled by a single enzyme. A 10-fold improvement in a flux through this hypothetical pathway can be achieved either through a 10-fold overexpression or through a 10-fold activity increase. In the latter case, pathway flux can be improved independent of gene overexpression and protein level and thus would impose a lower overall metabolic burden. As a result, controlling pathways through altering enzyme kinetics would have a greater impact on overall productivity. However, cellular environments present numerous complex interactions and thus catalytic activity may also be influenced through phenomena such as allostery. Forms of enzyme inhibition must be overcome to provide the greatest impact in large-scale bioreactor production of small molecules [35]. Thus, directed evolution and protein engineering can solve a number of flux control problems inherent in metabolic pathways as depicted in Figure 1-1.

Figure 1-1: Illustrations of ways in which directed enzyme evolution can improve metabolic flux.



Directed evolution can improve metabolic flux through: **A)** Increasing a rate-limiting enzyme's kinetic properties, **B)** Reducing the production of byproducts, **C)** Altering an enzyme's cofactor preference, or **D)** Removing product inhibition.

Directed evolution can help alleviate rate limiting steps by increasing the catalytic rate of enzymes. As an example, pentose catabolic pathways (in particular, xylose catabolism) are a rate limiting step in the utilization of lignocellulose by the yeast *S. cerevisiae*. A xylose isomerase-based heterologous pathway demonstrated limited success and directed evolution was used to improve the catalytic rate and hence pathway functions. Specifically, a mutant xylose isomerase was identified with a 77% improvement in V_{max} *in vitro*. Overexpression of this mutant in *S. cerevisiae* increased the growth rate in xylose by 61-fold and the production of ethanol by 8-fold when combined with the overexpression of a xylulokinase enzyme [17]. In another example,

directed evolution has been applied to overcome a bottleneck present in carotenoid biosynthesis in *E. coli*. Specifically, geranylgeranyl diphosphate (GGPP) synthase (GAS) from *Archaeoglobus fulgidus* was subjected to directed evolution and random mutagenesis libraries were screened to isolate a variant that increased lycopene production by 100% [21]. These examples demonstrate that metabolic flux control of rate-limiting steps can be alleviated through directed evolution as demonstrated in Figure 1-1a.

Beyond enhancing native catalytic function, directed evolution is a powerful tool for introducing novel catalytic functions in promiscuous enzymes. Examples include the isolation of a cytochrome P450 monooxygenase mutant capable of oxidizing octane with a k_{cat}/K_m of 3.5 x 10^4 $M^{-1}s^{-1}$ compared to the k_{cat}/K_m of 6.0 x 10^7 $M^{-1}s^{-1}$ the enzyme possesses towards its optimal substrate [16]. Another directed evolution study of the same enzyme sought to produce a mutant capable of performing olefin cyclopropanation, an important reaction in organic synthesis that is not present in a biological context. Mutants isolated from this study were able to provide yields greater than 30% with strong diastereomer preferences (>80:1) [36]. Additionally, a galactose oxidase with glucose 6-oxidase activity was isolated using only one round of mutagenesis [22]. Similarly, the metal binding site of the paraoxonase-1 enzyme was restructured using mutagenesis to reduce lactonase activity (100-600-fold) while improving organophosphate hydrolase activity (~300-fold for parathiol) [37]. In another example, DNA shuffling was employed to improve 3-isopropylmalate dehydrogenase activity 65-fold for a bacterial homoisocitrate dehydrogenase [38]. This approach has the potential to be extended to various attractive metabolic engineering targets such as alcohol dehydrogenases, zinc finger transcription factors, cytochromes, and hydrogenases.

Significant product diversification via directed evolution is exemplified by engineering the various carotenoid enzymes to create pathways for producing novel C_{40}, C_{45}, and C_{50} carotenoids in *E. coli* [39, 40]. One study sought to diversify the variety of available extender units for polyketide synthesis by using directed evolution to increase the promiscuity of a malonyl-CoA synthetase. This study produced several variants that were able to produce acyl-CoAs using 9 different malonate analogues with activities comparable to the native enzyme [23]. From these advances, polyketide synthases that can recognize non-natural extender units may be used to increase the diversity of producible polyketides. In addition, directed evolution can be used to alter the specificity of promiscuous enzymes (Fig 1-1b), such as γ-humulene synthase, which can natively produce 52 different sesquesterpines with low specificity. Saturation mutagenesis yielded several mutants which not only improved both the total yield and the specificity for different products, but also substantially increased the titers for Z,E-α-farnesene and α-ylangene [41].

Beyond catalytic rate, an important property of enzymes is their dependency on cofactors, especially in the context of metabolic engineering. Traditional methods for cofactor engineering have proven beneficial in overcoming this problem; for example, the overexpression of an NADH Oxidase from *Streptococcus pneumonia* in *Klebsilla pneumonia* improved the glucose flux to Acetoin (mol/mol glucose) by 189% [42]. As an alternative approach, directed evolution can be used to change the cofactor dependencies of enzymes, as shown in Figure 1-1c. For the case of isobutanol production, a ketol-acid reductoisomerase mutant was identified that improves the enzyme's relative catalytic efficiency for NADH/NADPH from 0.003 to 185 (~6.2×10^4-fold) without sacrificing enzymatic activity. Overexpression of this mutant with an active NADH-dependent alcohol dehydrogenase improved isobutanol productivity from

1g/L to 13.4 g/L and at 100% theoretical yield whereas the expression of a transhydrogenase to restore NADPH only improved productivity 8.5-fold [24]. In another example relevant to metabolic engineering, a mutated acetohydroxy acid isomeroreductase that prefers NADH to NADPH was introduced into *Corynebacterium glutamicum* to resolve the cofactor imbalance caused by valine synthesis. Overexpression of this mutant resolved the NAD+/NADH imbalance, consequently improving the organism's fitness and increasing valine productivity by nearly 40% [43]. In another study, the cofactor requirement for a Cytochrome P450 monooxygenase was changed, using directed evolution, from NADH to hydrogen peroxide, which is drastically different in size and structure. This enzyme was subsequently utilized to improve the hydroxylation of naphthalene by 20-fold in *E. coli* by taking advantage of the 'peroxide shunt' [25].

A final aspect of enzymes that may be altered through directed evolution is allosteric interactions and feedback inhibition, illustrated in Figure 1-1d. The importance of addressing metabolite inhibition in metabolic engineering was demonstrated in the evolution of a tomato peel 4-coumarate:CoA ligase for enhanced phenylpropanoid production. In this study, two of six highly active variants isolated from fifty thousand colonies were shown to possess mutations that either greatly diminished or eliminated feedback inhibition [44]. Another study which applied directed evolution to address such regulation eliminated the feedback inhibition of chorismate mutase-prephenate dehydratase (CM-PDT) in order to increase production of L-phenylalanine in *E. coli*. A truncated version of CM-PDT insensitive to phenylalanine inhibition but with a 3-fold lower catalytic efficiency was subjected to directed evolution in order to improve activity. The resulting variant, when expressed with other genetic modifications, increased the phenylalanine yield from glucose to 60% of the theoretical maximum,

which was the highest yield coefficient reported [45]. In another example, citramalate synthase (*CimA*), which converts pyruvate and acetyl coenzyme A (CoA) into 2-ketobutyrate, was introduced into *E. coli* to enhance n-propanol and n-butanol production. While employing directed evolution to improve the thermophilic enzyme's activity at lower temperatures, researchers isolated a variant which also eliminated the wild-type enzyme's sensitivity to L-isoleucine inhibition. When this variant was overexpressed with additional pathway enzymes, production of n-propanol and n-butanol were subsequently increased by 9.2-Fold and 21.9-fold respectively [46]. These results highlight yet another strategy for increasing yields in metabolically engineered organisms through directed evolution to remove regulatory inhibition while retaining activity.

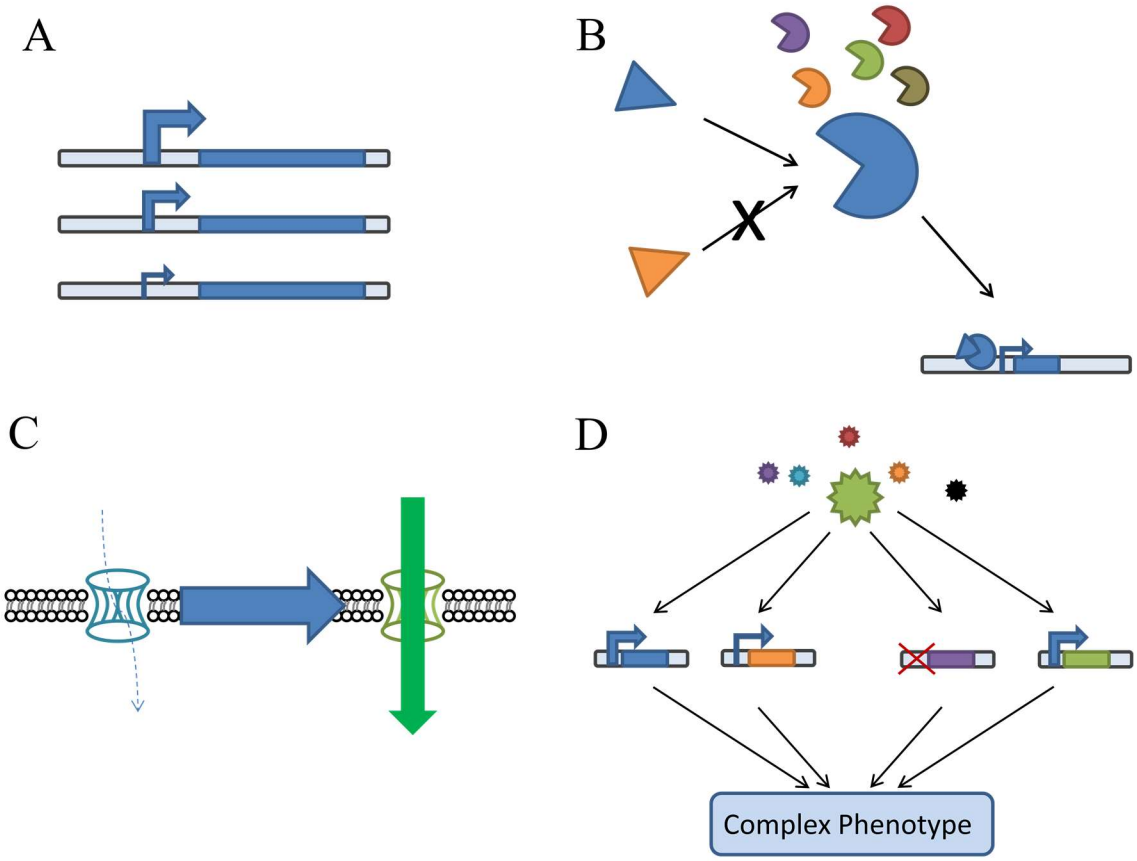
Collectively, these examples demonstrate that directed evolution of enzyme function can provide a means for fine-tuning alterations of cellular metabolism to address problems that gene knockouts and overexpressions cannot readily solve. While challenges still remain in selection and differences between *in vitro* and *in vivo* function, it is clear that directed evolution of enzymes is a potent approach for metabolic engineering.

1.1.2 Directed Evolution of Non-Catalytic Components of a Cell

While enzymes conduct the essential function of a metabolic pathway, additional non-catalytic components such as regulatory elements, binding proteins, transcription factors, and transporter proteins serve a critical role in defining cellular function. In each of these cases, directed evolution techniques may be used to improve function and hence utility in metabolic engineering applications, as illustrated in Figure 1-2. We highlight

several advances in metabolic engineering in the context of applying directed evolution to these non-catalytic components of a cell.

Figure 1-2: Illustrations of several approaches in which directed evolution of non-catalytic components can improve metabolic flux.



Directed evolution can improve non-catalytic aspects in cells through: **A**) Creating libraries of promoters of various strengths can allow precise modulation of expression and increase overall productivity, **B)** Transcriptional controllers or biosensors can be evolved to bind with high specificity and activate transcription of a target gene, **C)** Molecular transporter proteins can be evolved to increase flux of a desired substrate into a cell and/or products and toxins out, or **D)** Evolving global transcription factors can alter the expression levels of many genes simultaneously, leading to complex phenotypes that could not be engineered rationally.

1.1.2.1 Evolving regulatory elements

Regulatory elements, especially promoters, are a critical component of any metabolic engineering effort, especially when considering the engineering of entire pathways. Previous studies have demonstrated that directed evolution of promoters, in an approach termed promoter engineering, is an effective manner for diversifying promoter activity, as depicted in Figure 1-2a. This approach was first demonstrated by mutagenizing the constitutive bacteriophage promoter PL-λ and demonstrating a 200-fold range in expression. This library was then used to modulate expression of the *E. coli* gene *dxs*, which encodes an enzyme in the isoprenoid pathway. These results uncovered an optimum promoter strength for maximized lycopene production whereby stronger promoters were actually detrimental to productivity [47].

These results highlight the importance of modulating gene expression while engineering cellular metabolism. Constitutive overexpressions and gene knockouts are the two most commonly used genetic modifications, but these represent two polar extremes of a continuum, whereas the optimal level of expression is almost certainly somewhere between the two, requiring tunable levels of expression [48]. This is especially true when a knockout of a gene is lethal, but a "knockdown" could beneficially influence productivity. This concept was demonstrated using the *S. cerevisiae* TEF promoter where a mutagenesis approach resulted in a library of yeast promoters with a wide dynamic range. Variants from this TEF library were then used to modulate expression of GPD1 in *S. cerevisiae*. By replacing the native GPD1 promoter with the mutant TEF promoter library, the specific level of control GPD1 exerts on glycerol flux was quantified. Furthermore, the integration of a stronger mutant TEF increased glycerol production without the negative effect on growth rate observed in multicopy GPD1

overexpression, validating the idea that precise modulation can be useful in metabolic engineering efforts [49].

A similar approach was applied to the oxygen-responsive *S. cerevisiae* DAN1 promoter. Random mutagenesis created two variants that could be induced under less-stringent anaerobiosis than the wild-type promoter, which would enable gene expression to be induced in yeast fermentations simply by oxygen depletion. Furthermore, both variants were stronger than the native DAN1, with higher maximal expression [50].

In addition to engineering inducible promoters, constitutive regulatory elements can also be created or altered with this approach. One such study utilized four rounds of random mutagenesis and selection to generate an *E. coli* promoter approximately 2-Fold stronger than the commonly used tac promoter. Interestingly, randomly selected eukaryotic DNA was used as a template, indicating that it may be relatively easy to generate strong constitutive promoter activity in prokaryotes [51].

Proteins may influence metabolic pathways by serving as cis-acting regulators and likewise can be altered via directed evolution. For example, the XylS transcription factor in *E. coli* activates the inducible *Pm* promoter in the presence of benzoic acid. Random mutagenesis of the *xylS* gene followed by gene shuffling resulted in a XylS variant that demonstrated 10-fold higher expression under induction [52]. Similarly, the AraC protein in *E. coli* natively activates the *ara* operon only in the presence of L-arabinose. By using both saturated mutagenesis of amino acids in the arabinose binding site and global error-prone PCR, a variant was isolated that activates transcription in the presence of D-arabinose with a sensitivity to concentrations as low as 0.1 mM, while displaying no activity in up to 100 mM of L-arabinose [28]. Another study constructed a similar library of araC mutants, and isolated a variant that was significantly less sensitive to inhibition by IPTG, thus eliminating the possibility of “cross-talk” [53]. This new

construct will allow dual-control of pathways through multiple inducers. These studies collectively demonstrate the utility of directed evolution of inducible regulatory proteins to modify activation or inhibition of expression in specific ways.

A similar example applying directed evolution to cis-regulators is the evolution of tyrR, a prokaryotic tyrosine-inducible transcription factor that activates *tp1* in *E. herbicola*, encoding for an enzyme used in the production of L-dopa. Using random mutagenesis, a variant of tyrR was identified that demonstrated 30-fold higher expression levels of Tp1 without the addition of tyrosine—a component that ruins purification [54]. This example demonstrates how directed evolution can be used to eliminate the need for inducible expression.

Another important tool for developing complex phenotypes is global Transcription Machinery Engineering (gTME), which applies directed evolution strategies to a transcription factor regulating the expression of many genes, as depicted in Figure 1-2d. In this approach, the expression levels of many genes are altered simultaneously, allowing the combinatorial optimization of a desired phenotype. By applying this concept to *spt15*, the gene encoding TATA-box binding protein, a strain of *S. cerevisiae* was evolved for ethanol tolerance by serially culturing cells containing mutated *spt15* genes in 6% ethanol. A variant was isolated that conferred significantly higher tolerance, resulting in 70% increase in specific productivity of ethanol, close to the theoretical maximum [27]. A similar approach has been demonstrated across organisms such as *E. coli* [55-57], *L. planarum* [58], and even in the plant species *Arabidopsis thaliana*, where one study used directed evolution of the plant transcription factor ERF to isolate a variant which significantly improved cold tolerance in transgenic plants, an important factor for growth and production [59]. Simultaneous changes to multiple gene expression levels would be impossible to predict or rationally explore. However, directed

evolution of regulatory machinery can achieve such complex phenotypes and is therefore a potent tool for metabolic engineering.

1.1.2.2 Evolving binding optimization

Directed evolution has been proven effective at modifying binding affinity. Such activity can be essential for metabolic engineering, as interactions between two proteins or between proteins and small molecules can be engineered to provide a desired response. For example, engineered interactions can be used to create sensitive and accurate biosensors that can detect changes in metabolite concentrations. Shifts in metabolic flux can also be actuated upon the presence of a hormone or metabolite using this approach, by the establishment of molecular and genetic "switches". Thus, binding affinity is an important property for synthetic engineering of organisms, as illustrated in Figure 1-2b.

One crucial application of the evolution of binding properties is developing sensitive and accurate biosensors. For example, one study increased production of lycopene in *E. coli* by engineering the *Ntr* regulon to detect intracellular concentration of a metabolite, acetyl phosphate, and activate enzyme pathways [60]. In a more complex biosensor system, metabolite-sensitive transcription factors can be utilized to maximize productivity by real-time adjustment of enzyme levels based on signals such as intermediate accumulation or nutrient availability [61]. To achieve this system, however, it is necessary to engineer proteins with specific binding properties, and directed evolution is an essential tool in creating these systems.

One approach to creating accurate biosensors through directed evolution is to detect binding affinity using fluorescence as a reporter, which can then be used as a selection screen. An example of this approach was demonstrated by linking an alkane-

activated transcription factor (AlkS) to GFP expression, then generating a random library of AlkS mutants through error prone PCR. Variants were identified that enabled a 5-fold improvement in detection signal in the presence of hexane [62]. Similarly, the LuxR transcription factor in *V. fischeri* was mutagenized to create variants that activated in the presence of a broad range of straight-chain acyl-HSLs, but not in the presence of the native signaling factor, 3-oxo-hexanoyl-homoserine lactone (3OC6-HSL). The variants were screened using a dual selection system consisting of alternating rounds of selective pressure. Ultimately, this method resulted in a mutant transcription factor with over 50,000-fold change in specificity compared to the wild-type [63].

1.1.2.3 Evolution of transporter proteins

A final non-catalytic, yet essential, metabolic function is the transport of small molecules into and out of the cell. This often overlooked step in metabolic pathways inherently limits the maximal potential productivity of a small molecule by reducing substrate availability or constricting product secretion. Flux control exerted by substrate transport is especially problematic when engineering cells to catabolize less natural substrates, such as the pentose sugars that make up a large proportion of lignocellulosic biomass. Consequently, efficient transport and utilization of exogenous sugars is a limiting factor in the conversion of biomass to fuels or chemicals in *S. cerevisiae* [64]. To address this problem, several recent efforts have applied directed evolution to transporter proteins to significantly improve productivity, shown in Figure 1-2c.

One such effort employed directed evolution to improve heterologous xylose transporters in *S. cerevisiae*. Major facilitator superfamily transporters were subjected to error prone PCR and improved mutant transporters were identified by selecting the

largest colonies on xylose plates. In this study, the best variant improved the growth rate by 70% when cultured in xylose, and in vitro enzymatic assays confirmed that the V_{max} of the mutant transporter was nearly 50% higher. These results suggest that applying directed evolution to engineer molecular transporters can alone improve the overall growth rate and catabolic rate using non-native substrates [65].

1.2. Current Limitations of Evolutionary Engineering

While many successful examples exist, future efforts to apply directed evolution to metabolic engineering must rely on engineering entire pathways together—in essence, a metabolic pathway protein engineering approach. This simultaneous optimization will allow for synergies between pathway steps to be discovered and enhanced. Side-reactions, flux of intermediates to other pathways, and co-factor dependencies are all difficult or impossible to consider when engineering only a single rate-limiting step. Yet, the challenges of pathway protein engineering are exponentially more difficult than that of engineering a single enzyme.

1.2.1 Library Size Limitations

Although directed evolution has proven successful in its ability to improve and alter enzyme activity, the larger endeavor of pathway engineering requires an extremely large library of mutations. In engineering an entire pathway, the library of variants required to combinatorially explore the interactions and synergies between enzymes is orders of magnitude higher. Currently, traditional *in vitro* mutagenesis is encumbered by intrinsic limitations of host transformation efficiency, especially in more industrially- and medically-relevant eukaryotic systems. Next-generation evolution techniques aim to

accelerate the discovery of improved variants through continuous rounds of mutagenesis/selection on specific DNA cargo with reduced costs using *in vivo* diversity generation [66]. It has been demonstrated that mutational throughput can be increased in *E. coli* in an *in vivo* continuous process using phage, enabling the rapid evolution of parts [67-70]. However, this approach is best suited for phenotypes linkable to phage growth (e.g. DNA-binding proteins) and cannot be applied to eukaryotes. No current methods are well-suited for the deep meso-scale optimization (i.e. generation of all single-nucleotide substitutions to multi-kb pathways and gene networks) necessary for evolution of complex multi-part systems, or continuous evolution in eukaryotes.

1.2.2 Pathway Assembly Limitations

In order to effectively engineer pathways of enzymes, novel tools will also need to be developed for the efficient simultaneous expression of multiple proteins. Commonly, heterologous pathways constructed in yeast use a promoter and terminator motif for each individual gene, often using multiple vectors. This can be highly disadvantageous for several reasons, including lack of multiple well-characterized regulatory elements available, genomic instability due to recombination, and the cost and difficulty in synthesizing non-coding DNA.

For evolutionary approaches, this multiple vector approach is especially problematic. In order to simultaneously mutagenize multiple genes, they must be assembled in a single cassette. With each gene requiring its own promoter, this results in mutagenesis of large stretches of non-coding DNA, reducing the frequency of coding mutations proportionally. For directed evolution, it is highly preferable to minimize the entire pathway sequence in order to increase the likelihood of generating beneficial

mutations. Alternatively, each gene can be mutagenized independently and assembled prior to transformation [71], although this introduces the potential of bias during each cloning step.

One promising approach is the utilization of viral 2A sequences. These short sequences are about 20 amino acids in length and are found encoded in viral genomes to allow translation of single large polypeptides, which self-cleave into multiple smaller proteins. When a ribosome translates a 2A sequence, the nascent polypeptide quickly cleaves to form a separate protein, while translation continues downstream to form a second protein [72]. In this way, 2A sequences allow a single promoter to drive expression of an open reading frame encoding multiple proteins. Furthermore, such sequences have been previously shown to be active in yeast, making them a potentially useful tool in the evolution of multi-gene pathways [73, 74].

1.2.3 High-throughput Screening Limitations

Metabolic pathway engineering approaches are unquestionably still limited by our current screening capacity. Screens for improved enzymatic activity performed in vitro (also via cell free biological systems) may be able to screen large libraries of enzyme quickly, but are subject to a large number of false positives that do not necessarily function *in vivo*. Traditional techniques such as GC, HPLC, and MS can provide excellent data allowing for the selection of mutants improving product generation; however, they are low-throughput when compared to *in vitro* assays [20]. The rapid use of aptamers and other binding proteins linked to a fluorescent protein can enable higher-throughput techniques such as FACS to be used. In addition, approaches coupling cell growth to the specific production of the desired metabolite have demonstrated success

without the need for additional screening. For example, one approach coupled cell growth with succinate production in *E. coli* by knocking out genes encoding other NADH oxidizing pathways; in this strain, growth rate alone selected for mutations beneficial to succinate production without additional selection or screening [75]. Furthermore, modeling can guide this approach to maximize desirable evolutionary outcomes [76-78]. However, in order for directed evolution to fully benefit metabolic engineering, screening technology needs to be improved significantly to better match our ability to create sequence diversity.

This is especially relevant for production phenotypes, which, unlike growth phenotypes, are significantly more challenging to screen in a high-throughput manner. Specifically, it is usually difficult and slow to accurately measure the quantity of the desired product which an isolated strain produces. Although a relatively large library of variants can be created with traditional methods, only a small fraction can be tested in a reasonable time frame. Instead, growth-based selection is often used, by placing the library of variants in conditions such that desired phenotypes will naturally grow faster. However, this has been difficult or impossible to do for most production phenotypes because production of a desired product is often antithetical to growth.

Several recent advances in screening and mutation technologies demonstrate how such challenges will be met in the future. One approach for accurate and high-throughput screening being developed harnesses the power of microfluidics, where thousands of picoliter-sized drops can be analyzed per second. This miniaturization may represent the future of high-throughput screening, since significantly less time and material is needed. One such approach was used in the directed evolution of horseradish peroxidase enzyme, in which over 10^8 individual variants were screened in only 10 hours. It is estimated that such technologies already represent 1000-fold increases in speed and over 1-million-fold

reductions in cost [79]. When coupled with new analytic capacities, these microfluidic approaches may enable the future of metabolic pathway protein engineering as well as high-throughput synthetic biology [80].

Taken together, it is quite clear that as screening and mutation technology continues to improve, directed evolution will certainly play an ever greater role in the metabolic engineering of microorganisms. Protein engineering can solve many of the problems of metabolic engineering in a more efficient, fine-tuned manner. As techniques continue to be refined by increasing variant library size and enabling more accurate high-throughput screening, increasingly complex and less understood systems can be studied. Directed evolution of pathways will enable the next generation of metabolic engineering and industrial production in microorganisms.

The work described in the remainder of this document represents a significant contribution to the development of these techniques, and will address some of the most significant limitations discussed above. Chapter 2 describes a novel approach for *in vivo* Continuous Evolution (ICE), which allows the rapid creation of unprecedentedly large libraries in yeast. Chapter 3 discusses further characterization and investigation of this technology, and explores ways it could be improved or expanded upon. Chapter 4 delves into a new method for the creation of RNA biosensors that could enable high-throughput screening, in order to effectively search these large libraries. In Chapter 5, these new biosensors are applied in microfluidic systems, in an approach called RNA Aptamer In Droplet Screening, or RAPID Screening, to demonstrate how such a high-throughput screening method can be used in a relevant context. In Chapter 6, viral 2A sequences are characterized and developed to allow poly-protein expression from a single genetic cassette, which can be essential for the evolution of entire pathways. Finally, in Chapter 7 the conclusions and future directions of this research are explored, and it is discussed

how the novel approaches developed and validated here will make a significant contribution to the fields of directed evolution and metabolic engineering and to industrial production in microorganisms.

Chapter 2: *in vivo* Continuous Evolution of Genes and Pathways in Yeast [2]

2.1 Chapter Summary

Directed evolution remains a powerful, highly generalizable approach for improving the performance of biological systems. However, implementations in eukaryotes rely either on in vitro diversity generation or limited mutational capacities. Here, we synthetically optimize the retrotransposon Ty1 to enable *in vivo* generation of mutant libraries up to $1.6*10^7$ L^{-1} $round^{-1}$, which is the highest of any *in vivo* mutational generation approach in yeast. We demonstrate this approach by utilizing *in vivo*-generated libraries to evolve single enzymes, global transcriptional regulators, and multi-gene pathways. When coupled to growth selection, this approach enables in vivo continuous evolution (ICE) of genes and pathways. Through a head-to-head comparison, we find that ICE libraries yield higher-performing variants faster than error-prone-PCR-derived libraries. Finally, we demonstrate transferability of ICE to divergent yeasts, including *Kluyveromyces lactis* and alternative *S. cerevisiae* strains. Collectively, this work establishes a generic platform for rapid eukaryotic directed evolution across an array of target cargo.

2.2 Introduction

While directed evolution [81, 82] has generated solutions to engineering problems [27, 83], established novel functions [84, 85], and provided insights into evolution [68, 86], traditional *in vitro* mutagenesis techniques are encumbered by long and costly design-build-test cycles, restrictive requirements for hands-on manipulation of nucleic

[2] Adapted from Crook, N., et al., *In vivo continuous evolution of genes and pathways in yeast*. Nature Communications, 2016. 7: p. 13051. Nathan Crook, Joseph Abatemarco & Jie Sun all contributed equally to this work.

acids, and intrinsic limitations of host transformation efficiency. These limitations become especially poignant when attempting to optimize larger genetic systems (e.g. entire pathways including regulatory DNA), especially in more industrially- and medically-relevant eukaryotic systems. Indeed, the throughput of novel microfluidics-based screening technologies currently outpaces throughput for generation of genetic diversity in these systems [80]. Next-generation evolution techniques aim to accelerate the discovery of improved variants through continuous rounds of mutagenesis/selection on specific DNA cargo with reduced costs using in vivo diversity generation. It has been demonstrated that mutational throughput can be increased in *E. coli* in an *in vivo* continuous process using phage, enabling the rapid evolution of parts [67-70]. However, this approach is best suited for phenotypes linkable to phage growth (e.g. DNA-binding proteins) and cannot be applied to eukaryotes. Genome-editing technologies (such as MAGE [87] and CRISPR-Cas9 [88, 89] have enabled discovery of sequence-function relationships across a wide range of species but remain, at least at present, a method for introducing finite, defined mutations across several base pairs (ideal when important structural features are known) and are not well-suited for kilobase-scale directed evolution applications (which are necessary when structural features are largely unknown). In yeast, recent proof-of-concept demonstrations of continuous evolution (1) have suffered from low mutagenic rates and the necessity for mutant expression from weak promoters [90], or (2) require in vitro library generation [91]. None of these new methods are well-suited for evolution of multiple metabolic pathway genes simultaneously.

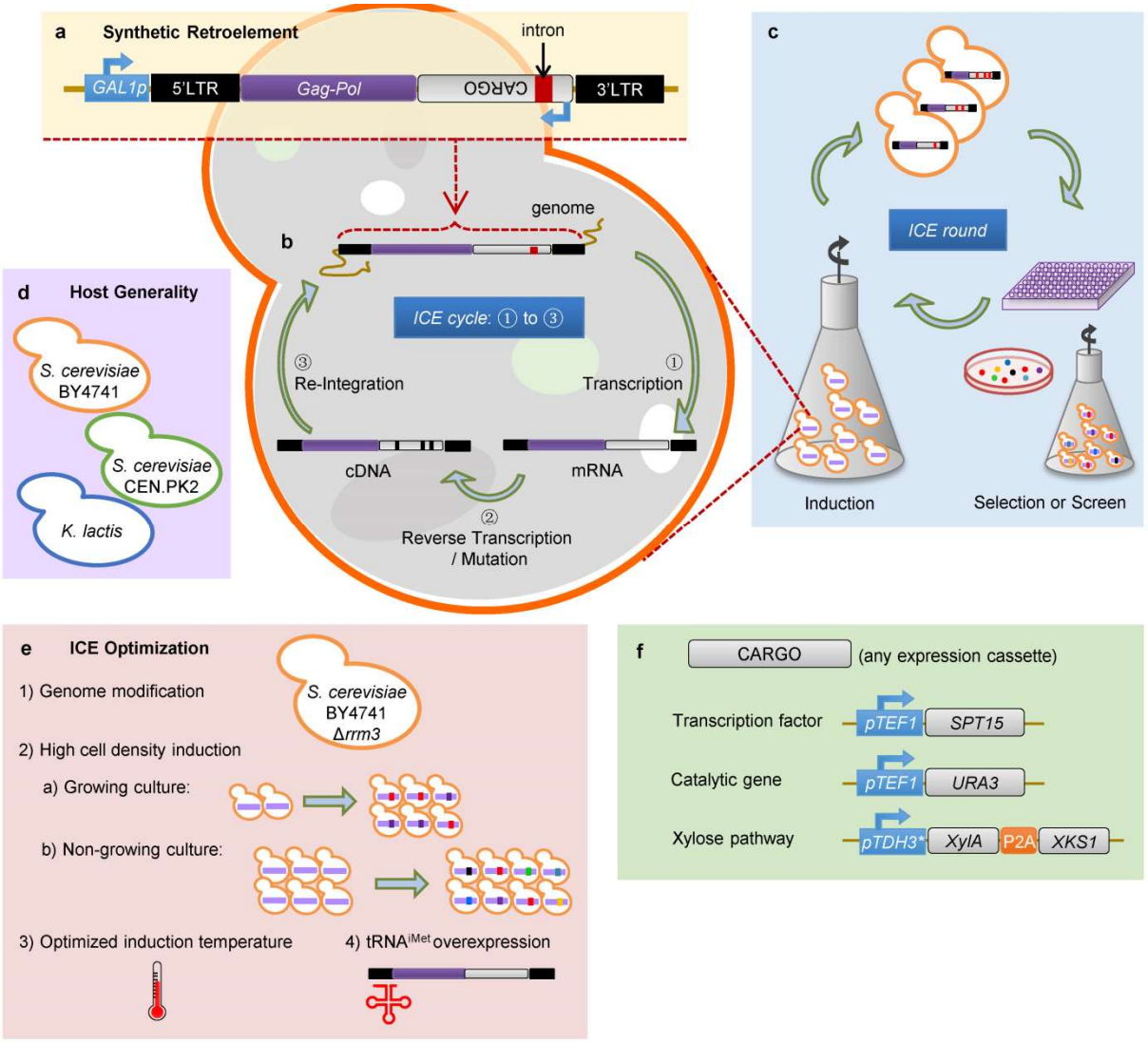
To fill this gap, a scalable, *in vivo* mutagenesis system was established in yeast by engineering its native retroelement Ty1 (Fig 2-1). The replication cycle of Ty1 proceeds via an RNA intermediate that is converted into cDNA through an encoded reverse

transcriptase [92]. Importantly, retrotransposons like Ty1 are found in all eukaryotic genomes, making this approach potentially generalizable to multiple cell hosts of interest [92]. Previous studies have demonstrated the potential for heterologous gene expression from Ty1 when inserted between Ty1RT and the 3'LTR [93, 94]. Thus, the error-prone nature of Ty1 replication [95, 96] coupled with the capacity for continuous retrotransposon cycling could enable a unique mechanism for *in vivo* directed mutagenesis of synthetic DNA (denoted here as "cargo") in eukaryotes a manner that is scalable with cell count (Fig 2-1a). In such a scheme, one cycle of *in vivo* mutagenesis is defined as the per-cell process of Ty1-cargo transcription, reverse transcription, and re-integration to a stable genetic context (Fig 2-1b). Since yeast cell densities can routinely exceed 10^{10} per liter (and even 10^{12} per liter in controlled fermentations), library size can easily exceed that of current *in vitro* techniques even with low mutation or transposition rates. A complete round of *in vivo* Continuous Evolution (ICE), in analogy to traditional directed evolution, is achieved at the culture level by allowing multiple cycles to occur through simple cell outgrowth, screening the resulting *in vivo* library, and isolating the best variant. As such, this approach can enable high-throughput, hands-off, scalable mutagenesis of desired parts and pathways (Fig 2-1c). For some applications, rounds may occur continuously and growth-associated phenotypes can be selected in tandem with mutagenesis, thus enabling ICE. In other applications, independent rounds may be desirable to segregate dominant mutations from background genetic drift.

By tuning the expression of key regulators of Ty1 transposition, we increase library size achievable using this system and confirm its ability to impart a useful error rate to an encoded cargo gene. Here, we apply this system to the directed evolution of a variety of synthetic parts, including single enzymes, regulatory factors, and multi-enzyme pathways, realizing substantial and significant improvements to performance in each

case. We further demonstrate that ICE enables the recovery of superior mutants more quickly than error-prone PCR. Finally, we show that ICE enables *in vivo* mutant generation across divergent strains of yeast, indicating its applicability toward a wide range of eukaryotic systems.

Figure 2-1: Rationale and schematic for in vivo continuous evolution (ICE) in yeast.



In the operational scheme of ICE: **A)** genetic cargo of interest is cloned into the genome of an inducible Ty1 retrotransposon; **B)** upon induction of retroelement transcription, the encoded reverse transcriptase is expressed, converts the Ty1 genome (including the cargo) into cDNA in an error-prone manner, and then this cassette is re-integrated into a stable genomic locus. This process is defined as one cycle; **C)** the procedure of inducing mutagenesis to a bulk culture and selecting for improved variants is analogously defined as one round. In this work, we **D)** apply this approach to several divergent strains and species of yeast, **E)** iteratively improve the efficiency of Ty1 retrotransposition through deletion of *rrm3*, reducing temperature, increasing cell density, and increasing expression of limiting cellular components, and **F)** apply this improved system to the evolution of transcriptional activators, single enzymes, and multi-enzyme pathways.

2.3 RESULTS

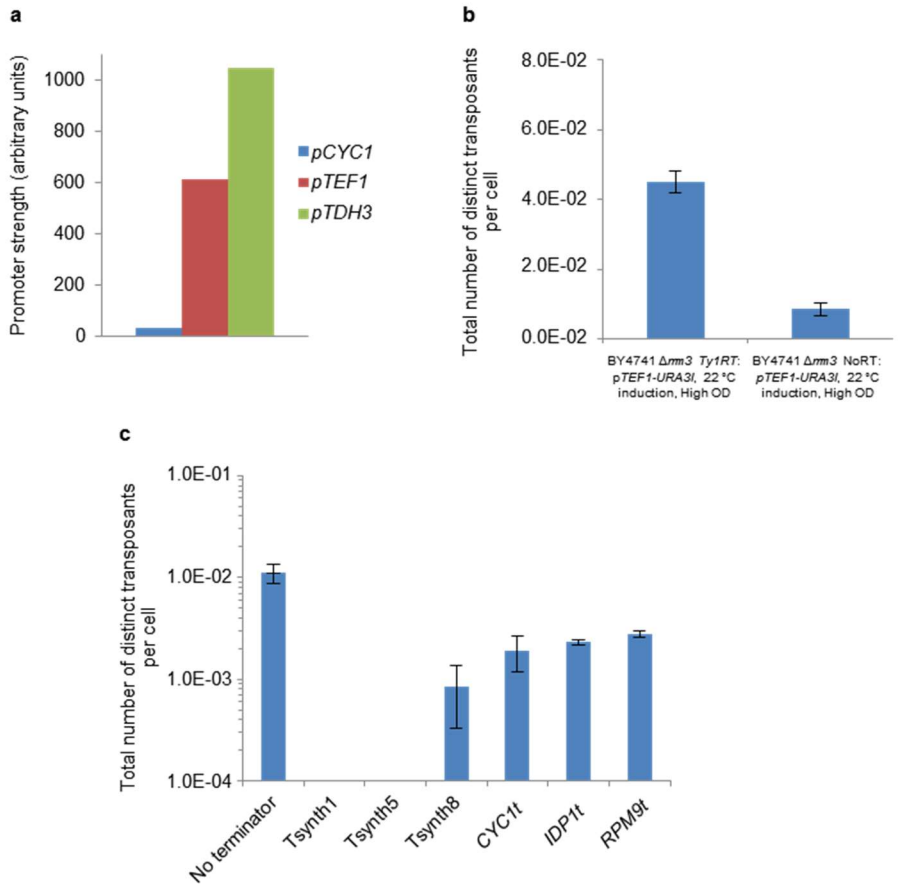
2.3.1 Engineering Increased Transposition Rate

After demonstrating basal functionality of the plasmid-based synthetic retroelement through induction at low cell density, we wished to develop strategies for increasing transposition rate of Ty1. To this end, we investigated cargo expression level, gene knockouts, cell density, induction temperature, and initiator methionine tRNA expression level as potential drivers of increased transposition rate.

2.3.1.1 Tuning Cargo Expression Increases Transposition Rate

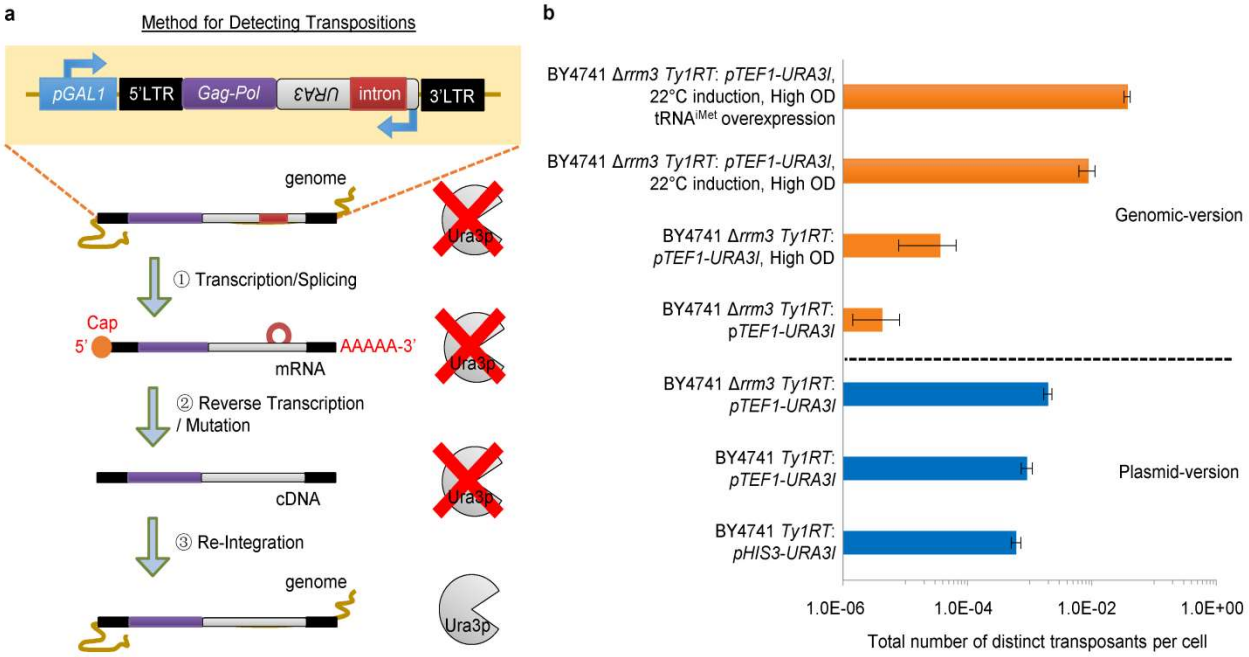
We first investigated the effect of cargo transcription rate on Ty1 transposition. Although strong promoters (such as *pTDH3*) are desirable for cargo overexpression, their high transcription rate may interfere with that of *pGAL1*, thus lowering transposition rate and library size. Out of 3 yeast promoters (*pCYC1*, *pTEF1*, and *pTDH3*, representing low, medium, and high transcriptional output, respectively [97] (Fig 2-2a)), we observed the highest transposition rate when the *pTEF1* drove expression of URA3 (Fig 2-3b and Fig 2-4a). We used this promoter in future benchmarking experiments.

Figure 2-2: Improvement of Ty1 Transposition.



A) Promoters used in this study. Indicated promoters were used to drive yellow fluorescent protein and fluorescence was measured using flow cytometry. **B)** Transposition of RT-containing and RT-less retroelements in a genomic context at 22ºC. **C)** The effect of terminators on transposition rate. No transpositions were observed for Tsynth1 or Tsynth5. Strains containing the appropriate retroelement were exposed to galactose at high OD for (**B**) and (**C**) for three days and then plated on uracil-deficient media to count transposants. Error bars in (**B**) and (**C**) represent the standard deviation of biological triplicates.

Figure 2-3: Iterative improvement of synthetic Ty1 transposition rate and scheme for detection of retrotransposition using an intron-marked, *URA3*-containing retroelement.



A) *URA3* is inserted into the retroelement in the reverse orientation relative to transcription from the *pGAL1* promoter. The presence of an intron in the same transcriptional direction as *pGAL1* prevents mRNA originating from the *URA3* promoter from being correctly spliced and initiating Ura3p synthesis. Upon transcription from *pGAL1*, the intron is spliced. This mRNA cannot give rise to Ura3p due to *URA3* being present in the reverse orientation on this transcript. However, once mRNA is converted into cDNA, a functional *URA3* expression cassette is formed, and integration of this cDNA into the genome ensures a heritable *URA3*$^{+}$ phenotype. **B)** Strain background, induction conditions, and expression of critical Ty1 components were modified to improve transposition rates of the synthetic retroelement. Error bars for the plasmid version represent 95% confidence intervals obtained via fluctuation analysis of biological triplicates, and error bars for the genomic version represent the standard deviation of biological triplicates.

2.3.1.2 rrm3 Deletion Increases Transposition Rate

Ty1 replication is known to be highly regulated by various host factors [98]. To evaluate a coupling between genotype and function, we performed an extensive literature

search and used the yeast haploid knockout collection to identify knockout phenotypes which enabled increased rates of Ty1 replication [98-100]. Of the various genotypes tested (Fig 2-4b), deletion of *rrm3* most significantly increased transposition rates in *S. cerevisiae* BY4741 (Fig 2-3b). Rrm3p plays a role in DNA repair, which may influence retrotransposition as it is dependent on homologous recombination [98]. Several other combinations of targets were evaluated, but these combinations did not exceed the transposition rate beyond that of Δ*rrm3* alone (Fig 2-4b). Therefore, *S. cerevisiae* BY4741 Δ*rrm3* was used for all subsequent experiments.

2.3.1.3 Genomic Integration of Ty1 Decreases Transposition

Subsequently, we elected to move the retroelement into the genome in order to gain a more accurate picture of retroelement behavior in its final context. We utilized BY4741 Δ*rrm3* with a genomically-integrated, *pTEF1*-containing retroelement at high OD, and determined that the transposition rate was significantly inhibited compared to the plasmid-based retroelement (Fig 2-3b).

2.3.1.4 Inducing Transposition at High Cell Density Increases Library Size

In all initial experiments, Ty1 transposition was induced when cells were at a low optical density and continued as cells divided. However, this growth can significantly reduce effective library sizes as mutations that occur early during growth can dominate the resulting culture during outgrowth [101]. We aimed to increase library sizes by inducing transposition for the same length of time (3 days), but at a much higher initial cell density (OD_{600}=1). In this condition, additional cell growth would have a greatly reduced effect on library size. This condition significantly increased the

retrotransposition rate (Fig 2-3b), and all subsequent inductions were carried out at high cell density.

2.3.1.5 Reducing Induction Temperature Increases Transposition

We next made use of the known temperature sensitivity of Ty1 [102] by inducing transposition at a lower temperature (22 ºC). This modification greatly improved transposition rate (Fig 2-3b). Interestingly, it also increased basal activation of our inducible Ty1 retroelement in the absence of a cis-encoded reverse transcriptase (Fig 2-2b), which could be due to the activation of endogenous Ty1 elements that are natively repressed at 30 ºC.

2.3.1.6 Increasing $tRNA^{iMet}$ Expression Increases Transposition

Based on real-time PCR experiments (Fig 2-4c and 2-4d), we noted that induction of Ty1 RNA levels by *pGAL1* greatly exceeded resulting cDNA levels produced by Ty1RT. As Ty1 replication is primed by the yeast initiator methionine tRNA ($tRNA^{iMet}$), we hypothesized that the concentration of this tRNA may be limiting transposition rates. By overexpressing tRNAiMet from several promoters, we observed greatly improved transposition rates (Fig 2-3b, Fig 2-2e). In particular, by overexpressing $tRNA^{iMet}$ using its native promoter and terminator on a high-copy plasmid, transposition rate could be improved by ~3.5-fold, and this increase was accompanied by a corresponding increase in cDNA levels (Fig 2-2f). All subsequent experiments used this overexpression strategy.

Figure 2-4: Improvement of Ty1 transposition.

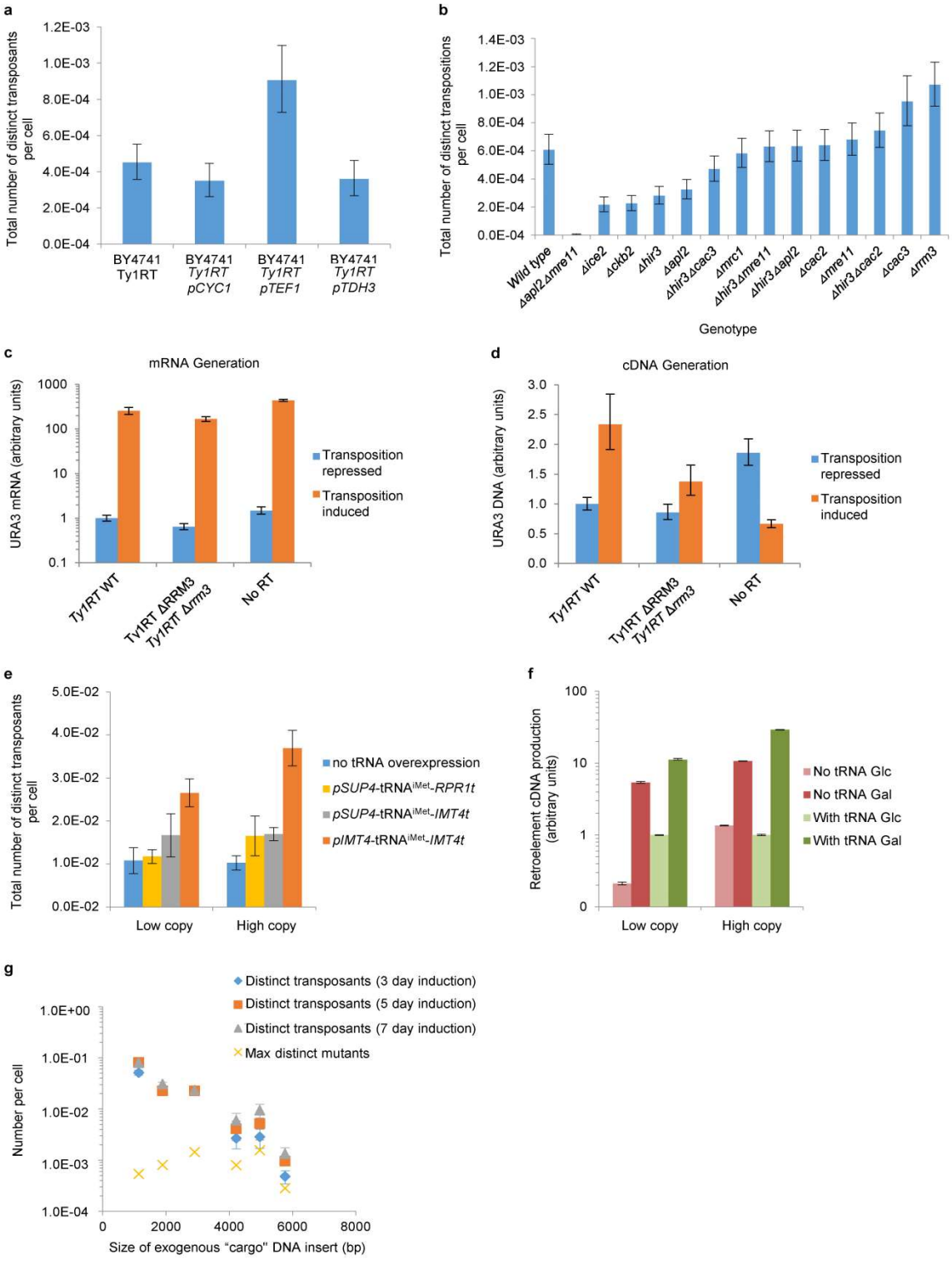


Figure 2-4 (continued): A) Substitution of alternative promoters in retroelement. "Distinct transposants" refers to the number of unique cells in which Ty1 underwent a full retrotransposition cycle at least once. This uniqueness explicitly excludes daughter cells arising from the original transposed variant. **B)** Transposition rates for BY4741 knockout strains. Transcript **C**) and cDNA levels **D**) of engineered Ty1 retroelements. **E)** Transposition rate of strains overexpressing the initiator methionine tRNA *IMT4.* **F)** $tRNA^{iMet}$ overexpression improves cDNA synthesis. Low copy and High copy data were collected on different days and hence are normalized to their respective "with tRNA Glc" values. **G)**, Transposition rate and mutation rate conferred by retroelements containing cargo of various sizes. "Max distinct mutants" refers to the maximum number of mutants attainable in a cargo of a particular length, given a 0.15 kb^{-1} mutation rate and the maximum number of distinct transposants attainable for a particular cargo size (maximum is calculated over 3, 5, and 7-day induction times). Strains containing the appropriate retroelement were exposed to galactose at high OD for (**G**) and low OD for (**A**) and (**B**) for three days and then plated on uracil-deficient media to count transposants. For (**C**) and (**D**), cells were exposed to the appropriate carbon source at high OD for three days. Total DNA and RNA was extracted after induction, and nucleic acid levels were quantified using qRT-PCR. For (**E**) and (**F**), strains containing a genomically-integrated retroelement were exposed to galactose at 22°C at high OD for three days. Error bars in (**A**) and (**B**) represent 95% confidence intervals from biological triplicates. Error bars in (**E**) and (**G**) represent the standard deviation of biological triplicates. For (**C**), (**D**), and (**F**) error bars represent the standard deviation of technical triplicates.

2.3.2 Characterization of ICE Transposition

2.3.2.1 Characterizing the Effect of Cargo Length on Transposition

Since it is highly desirable to include long sequences consisting of multi-gene pathways in the inducible retroelement, the effect of transcript length on transposition rate was characterized. Specifically, we inserted gene fragments as additional cargo between the URA3 reporter gene and the reverse transcriptase gene and then measured the resulting transposition rates. In addition, we measured transposition rates after 3, 5, and 7 days of high cell density induction in *S. cerevisiae* BY4741 *Δrrm3*. These experiments clearly revealed a negative correlation between cargo length and retrotransposition rate. However, lengthening the induction time from 3 to 7 days increased the number of retrotransposition events, especially for constructs containing the longest sequences (Fig 2-4g). Importantly, relatively high transposition rates were maintained within approximately an order of magnitude as cargo size increased to roughly 5 kb, indicating that Ty1 is capable of generating diversity to a multi-gene pathway. It should be noted that this experiment combined each pathway element on the same mutagenesis cassette. However, it is also possible to distribute several multi-gene mutagenesis cassettes across the genome to enable simultaneous evolution on multiple segments of longer cargo.

2.3.2.2 Characterizing the Effect of Terminators on Transposition

When expressing multi-gene pathways in yeast, it is common to include a promoter before each gene and a terminator afterward. When inserting a multi-gene pathway into the Ty1 mutagenesis cassette, however, a terminator with bidirectional activity can significantly affect transposition, since the entire retroelement must be

transcribed prior to reverse transcription. To characterize this effect, several native and synthetic terminators were inserted after the *URA3* reporter gene in the synthetic retroelement [103]. These experiments showed that including terminators inside the retroelement can lower the rate of transposition, with several terminators eliminating activity altogether (Fig 2-2c).

While we did identify several terminators that reduced transposition rate to within one order of magnitude, we instead opted to utilize 2A sites, which allow a single promoter to drive expression of a fusion peptide that then self-cleaves during translation [74]. This strategy allowed the evolution of multi-gene pathways, such as the xylose pathway evolved here, without including any terminators between genes. In addition, it allows multi-gene pathways to be expressed from a single promoter, reducing the length of DNA needed in the cargo and thus increasing the rate of transposition (Fig 2-4g). Including 2A sequences as opposed to terminators thus allows our approach to attain a significantly higher library size for multi-gene pathways through two mechanisms: it avoids terminators and it reduces cargo length by only requiring a single promoter. However, it should be noted that 2A sites are as of yet unoptimized for use in yeast, such that cleavage efficiency may not be 100% in all contexts and thus may pose an issue for pathways in which the generation of fusion proteins would be undesirable (see Chapter 6 for further discussion). For these cases, we recommend the integration of multiple distinct mutagenesis cassettes into the same strain to enable the simultaneous directed evolution of pathway components.

2.3.2.3 Measurement of Mutation Rate Enabled by ICE

Next, we undertook a mutation reversion experiment to investigate the error rate of this approach in comparison to random drift/mutation. To do so, a non-functional *KanMX* antibiotic marker was constructed by inserting an artificial stop codon (generating *dKanMX*). Cells containing either a genomically-integrated copy of *dKanMX* (as a control for random drift / adaptive evolution) or a copy integrated as cargo in the optimized Ty1 retroelement (the "mutagenesis cassette") were exposed to galactose and plated on G418-containing media. The number of G418-resistant colonies observed conservatively demonstrated a 20-fold higher reversion rate using in vivo mutagenesis over random drift (Fig 2-5a). Additionally, sequencing isolated colonies from this experiment demonstrated that out of 49 sequenced resistant colonies, 43 were found to have a mutation reverting the artificial stop codon, thus demonstrating in vivo-generated mutations as the mode of action.

In order to directly compare the rates of Ty1-induced mutations and background genetic drift, the *dKanMX* marker was constructed for use in a mutation reversion assay, containing two point mutations that together prevent functional activity. The first is T405A, which introduces an artificial stop codon at the 135th residue, in the middle of the proposed active site [104]. Interestingly, this mutation alone does not inactivate *KanMX*, and expression of this variant confers G418 resistance. The second mutation, T530A, is a missense variant resulting in $Trp^{177}Leu$. Together, these mutations inactivate *KanMX*, and do not allow growth in the presence of G418.

To separate the effect of background genetic drift absent the optimized Ty1 retroelement, *dKanMX* was either integrated directly into the genome (g-dKanMX) or incorporated into a Ty1 mutagenesis cassette, which was also integrated into the genome (ICE-dKanMX). Both strains were grown to stationary phase and exposed to galactose,

then plated on media containing G418. The rate of reversion mutations could then be measured by counting G418-resistant colonies. Genomic DNA from 39 colonies of both strains was extracted, a PCR designed to amplify position 135 of *dKanMX* was performed, and amplicons were sequenced via Sanger sequencing. In 34 of 39 resistant colonies sequenced, mutations at position 135 reverted the artificial stop codon to one of several non-stop codons, thus activating the gene and conferring a G418-resistant phenotype. We hypothesize that the 5 remaining colonies displaying a G418-resistant phenotype in the absence of a mutation at residue 135 may be due to a mutation elsewhere in *KanMX* (e.g. position 177), or a mutation elsewhere in the yeast genome (e.g. a mutation generating an amber suppressor tRNA). To this end, we repeated this experiment with an additional 10 colonies, this time amplifying and sequencing both mutated loci (position 135 and 177). Of these 10, 9 have mutations at position 135 that were previously observed. However, no other mutations in *dKanMX* were observed, including at position 177, in any of the 10 sequenced revertants. As stated above, the generation of an amber suppressor tRNA may be the cause of the remaining resistant isolate. We therefore conservatively conclude that at least 88% (43/49) of G418-resistant colonies appeared as a result of direct Ty1-mediated mutagenesis to dKanMX.

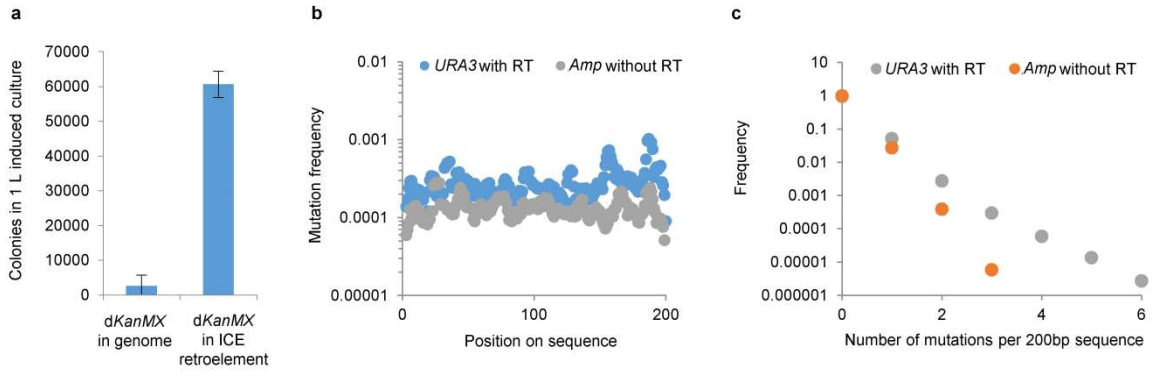
Next, we evaluated the full mutational rate and spectrum conferred during a single round of in vivo mutagenesis using next-generation sequencing. Yeast cells containing *URA3*-containing, Ty1 retrotransposons were induced in galactose for three days, after which intron-less *URA3* amplicons were generated via PCR of total DNA. As a negative control encompassing background genetic drift, PCR error rate, and sequencing error, a region of the Ampicillin resistance gene (*Amp*) of the same length (which is not reverse transcribed) was also amplified. These amplicons were then sequenced. Analysis of identified mutants showed a uniform error distribution across each amplicon, with *URA3*

consistently showing a higher mutation rate (0.28 kb-1) than *Amp* (0.13 kb-1) (Fig 2-5b). This 0.15 kb-1 increase in error rate above the combined effects of drift, PCR error, and sequencing error was thus due to Ty1 and was also reflected in an increased frequency of observing a given number of mutations per 200bp read in *URA3* versus *Amp* (Fig 2-5c). Finally, Ty1 exhibited a mutational spectrum commensurate with other commonly-used error-prone polymerases and displayed an error rate which is useful for directed evolution of genes and pathways (Table 2-1). Collectively, this analysis indicated that the Ty1 retrotransposon is a useful vehicle for introducing mutations to defined genes and pathways in vivo. This analysis that *URA3* was mutated at rates of 0.15 kb^{-1} (Fig 2-5b and 2-5c) generally agrees with previously-reported in vitro values for transcription/reverse transcription in Ty1 (0.05-0.25 kb^{-1}) [95, 96]. This mutation rate indicates that this system can effectively sample up to $1.6*10^7$ distinct mutants per liter (Fig 2-4g). This library potential is the highest reported for any *in vivo*, continuous directed evolution approach in yeast. This result demonstrates that the increase in mutation frequency achieved in this work is effective at evolving cargo of interest in a manner previously unexplored for eukaryotic systems.

The two methods (reversion mutation assay and next-generation sequencing) were also compared. With this established rate of 0.15 point mutations per bp, each transposition of the 717 bp *dKanMX* gene will result in an average of 0.10755 mutations. Given an approximate transposition rate of 10^8 per liter of culture (based on the length of the cargo cassette), 1.0 mL of culture plated on G418-containing media is expected to contain 10,755 mutations in *dKanMX*. Neglecting indel mutations, there are 8 mutations out of the 2151 possible that will eliminate the stop codon from *dKanMX*; assuming the mutations are distributed randomly and uniformly throughout the gene, this results in approximately 40 reversion mutations per 1.0 mL of culture that will be able to grow on

the G418 plate. This matches remarkably closely to the rate of colony formation observed (0.88*58 colonies per mL of culture= 51 per mL of culture), confirming the utility of this assay and the rate of mutation in ICE cargo.

Figure 2-5: Measurement of Ty1 Mutagenesis



A) Characterization of d*KanMX* reversion with or without ICE retroelement integrated in the genome. Cells were exposed to galactose and then plated on G418-containing media to count colonies. For (**A**), error bars represent standard deviation of technical triplicates. Mutation rates in Ty1 cDNA were determined through next-generation sequencing. Spatial distribution of mutation rates (**B**) and frequencies of observing a given number of mutations (**C**) of sequenced Ty1 cargo (*URA3*) as well as DNA not exposed to *Ty1RT* (*Amp*).

Table 2-1: Mutation spectrum of *Ty1RT*-RNAPII compared with other polymerases

	***Ty1RT* - RNAPII**	**Mutazyme II**	**Taq**	**DNAQ926**	**TP-DNAP1^{Y427A}**	**Yeast DNAP**
Bias Indicators						
Ts/Tv	1.06	0.9	0.8	1.4	0.17	0.98
AT->GC/GC->AT	0.28	0.6	1.9	0.3	0.44	0.48
A,T->N (%)	28.8	50.7	75.9	30	44.3	31.7
G,C->N (%)	69.4	43.8	19.6	70	55.7	68.3
Indels						
Insertions (%)	0.08	0.7	0.3	N.R.	N.R.	3.01
Deletions (%)	1.6	4.8	4.2			
Mutation Frequency						
Mutations per kb	0.15	3-16	4.9	0.05	0.00004	0.0000002

N.R. = Not Reported. DNAQ data taken from [67, 105]; TP-DNAP1^{Y427} data taken from[90]; Yeast DNAP data taken from [106]

2.3.2.4 Comparison of ICE to Error-Prone PCR

Finally, we sought to further compare the results obtained using *in vivo* derived mutagenesis with traditional *in vitro* mutagenesis. To this end, we generated two libraries of ura3 variants through error-prone PCR using standard approaches, resulting in libraries of two differing mutagenic rates each roughly 10^5 in size. These libraries were selected head-to-head with the *in vivo* mutagenesis libraries in uracil-deficient, 5-FOA-containing media with equivalent cell density trigger thresholds for subculturing (Fig 2-6a, see Methods). On the basis of bulk growth rate and subculturing frequency, it was clear that the in vivo mutagenesis derived libraries outperformed the error-prone libraries (reaching the threshold for subculturing 4 times versus 2 over the experimental timecourse) (Fig 2-5b). We next isolated 5 clones from each library/subculture and

compared their growth in 5-FOA concentrations ranging from 0 g L^{-1} to 1 g L^{-1} (Fig2-6c, Fig 2-7a,b). We found that the isolated strains derived from in vivo mutagenesis significantly ($p<0.05$, Mann-Whitney U test) outperformed those derived from in vitro mutagenesis in the majority (125/240) of possible comparisons (growth rate, maximum OD_{600}, and lag time). In 232/240 comparisons, in vivo mutagenesis was on par or better than traditional in vitro mutagenesis. This combined result is highly significant ($p<10^{-13}$, Fisher's method), demonstrating the utility of an in vivo mutagenesis approach. These two results indicated (1) that Ty1-generated libraries yield improved variants faster than error-prone PCR and (2) that Ty1-derived mutants significantly outperform those derived from error-prone PCR.

Figure 2-6: Comparison of ICE to error-prone PCR.

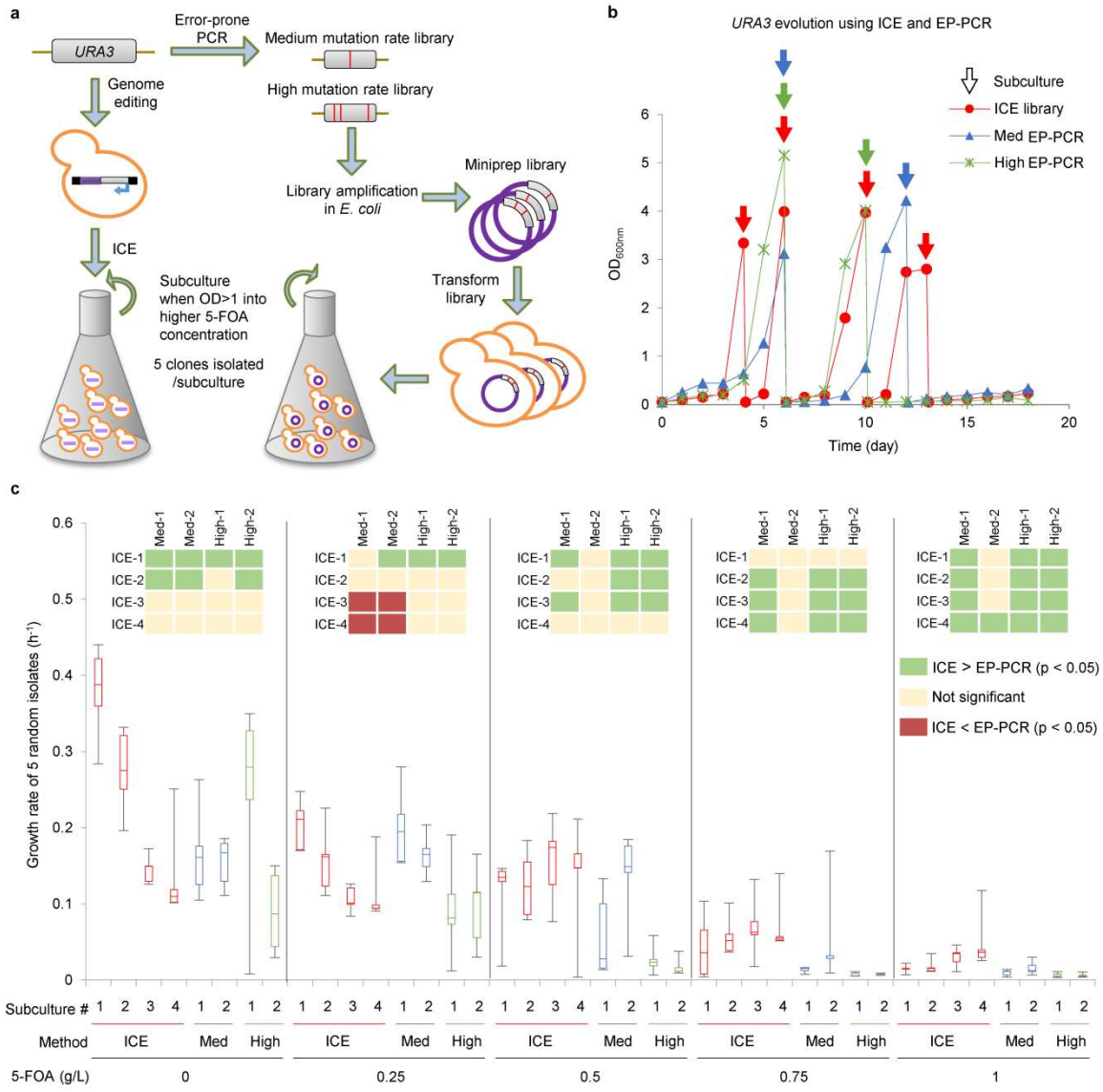
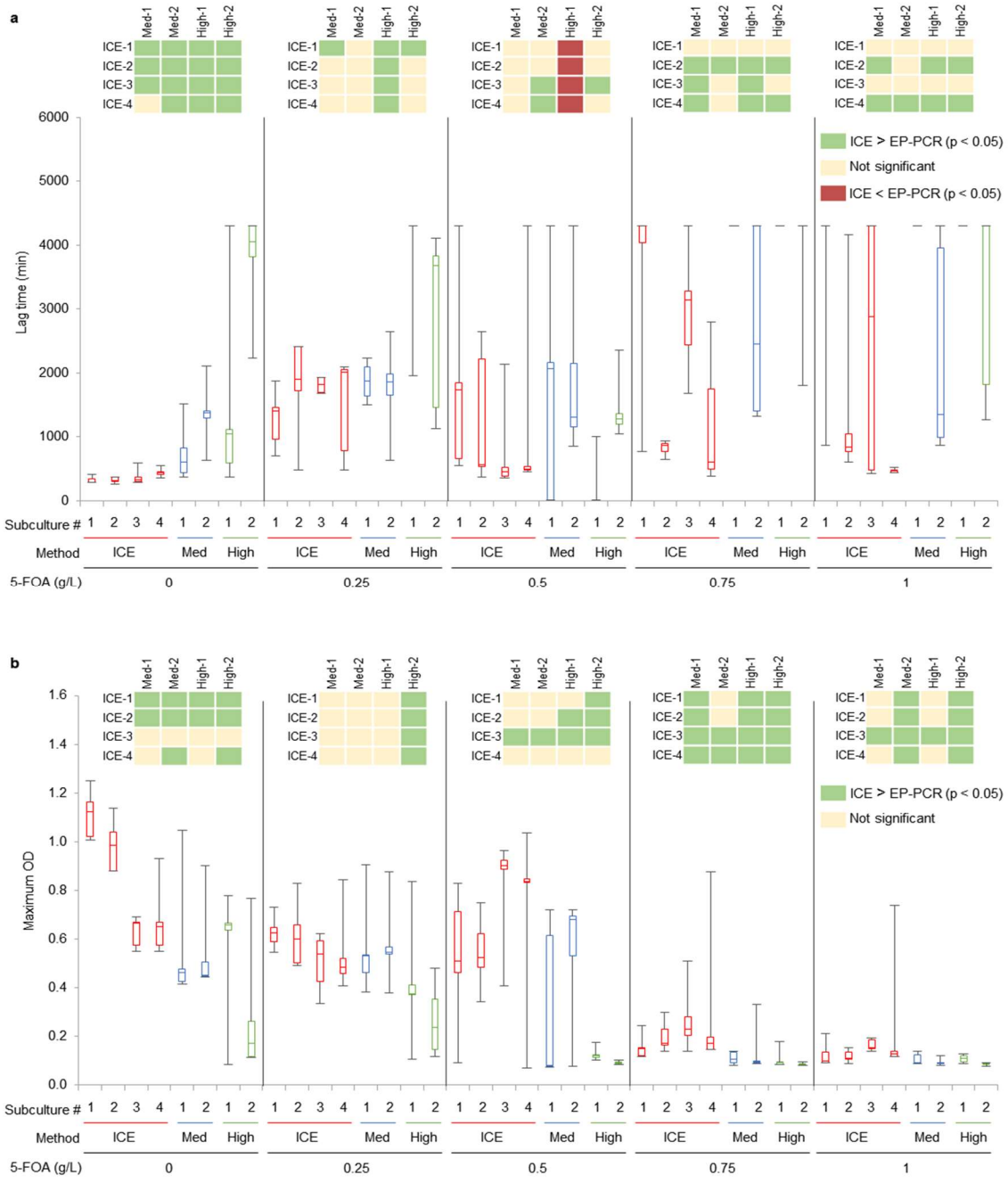


Figure 2-6 (continued): The ability of ICE to generate 5-FOA-resistant mutants of *URA3* was evaluated in parallel with error-prone PCR. **a**, workflow for parallel evolution experiment is diagrammed. **b**, culture cell densities were monitored as selections progressed. Each time the OD_{600} of the culture exceeded 3, a portion was transferred to a new flask containing a 0.25 g L^{-1} increased concentration of 5-FOA (denoted by the presence of a colored arrow). **c**, 5 random clones were isolated from each subculture (denoted by the number after the dash, eg. ICE-3 refers to clones isolated from the ICE library in subculture 3) and their growth was compared in varying concentrations of 5-FOA. Lines on boxplots in **c** represent, from top to bottom, maximum observed value, 75^{th} percentile, median, 25^{th} percentile, and minimum observed value. P-values used to generate heatmaps were computed using a Mann-Whitney U-test.

Figure 2-7: Comparison of Lag time and Max OD of ICE-derived and EP-PCR-derived mutants



A) Lag time. **B)** Max OD. Lines on boxplots represent, from top to bottom, maximum observed value, 75th percentile, median, 25th percentile, and minimum observed value. P-values used to generate heatmaps were computed using a Mann-Whitney U-test.

2.3.3 Applying ICE to Evolve Novel Phenotypes

With the *in vivo* mutagenesis system in place, we next investigated the utility of *in vivo*-generated libraries (along with different modes of selection) in the directed evolution of 3 broad classes of genetic cargo: small-molecule-converting enzymes, transcriptional regulators, and multi-enzyme pathways (Fig 2-1f). In these experiments, both regulatory regions and coding regions were subjected to mutation and selection (in contrast to common in vitro searches) to enable the evolutionary process to alter both expression and protein properties in its search for improved phenotypes, and to take advantage of the expanded library size afforded by this method. In each case, we utilized a genomically-integrated Ty1 element in a BY4741 Δ*rrm3* strain background, in which the cargo of interest was integrated between Ty1RT and the 3'LTR in the reverse orientation relative to Ty1 transcription and interrupted by an intron to facilitate mutant recovery. During evolution of *URA3* and *SPT15*, top mutants from each round were isolated and re-introduced to a fresh strain to eliminate any concurrent strain adaptation, while evolution of xylose catabolism proceeded in a continuous manner, demonstrating *in vivo* continuous evolution (ICE). The intent of these experiments was to investigate the breadth of evolvable cargo and the variety of experimental designs compatible with retrotransposon-generated libraries. It is clear that each of these experiments can serve as a launching point for further in-depth mechanistic analyses of mutant function as well as testing of different and subsequent evolutionary trajectories. As a mutagenesis technique, this approach is inherently phenotype-agnostic and could be further applied to additional targets which require alternative screening modalities.

2.3.3.1 Evolution of Improved Ura3p Substrate Specificity

Our first target involved engineering the substrate specificity of *URA3*, which encodes orotidine-5'-phosphate decarboxylase. Ura3p is an efficient catalyst which converts orotidine-5'-phosphate to uridine-5'-phosphate, yet also converts 5-fluoro-orotic acid (5-FOA) into 5-fluorouracil, a toxic compound. As a means to demonstrate the ability to modify enzyme promiscuity in a novel manner, we sought to isolate ura3 variants with a decreased ability to convert 5-FOA to 5-fluorouracil while maintaining their ability to enable uracil biosynthesis. Such variants have not, to our knowledge, been reported. To improve enzyme specificity for orotidine-5'-phosphate, we undertook simultaneous induction of mutagenesis and selection in uracil-deficient, 5-FOA-containing media, after which top variants were re-transformed into a fresh strain to exclude any adaptive, genomic mutations (see Methods). After two rounds, we isolated the best mutant, *ura3(3-5-2),* which significantly outperformed wild-type ($p<10^{-5}$, Fisher's method), and which conferred a 2.5-fold increased IC50 on 5-FOA to a fresh strain relative to a strain expressing the wild-type enzyme (Fig 2-9a). This variant contained two coding mutations ($Arg^{145}Ile$ and $Arg^{186}Lys$).

The first-round mutant *ura3(3-5)* contained a single coding mutation ($Arg^{145}Ile$) that resides on an outer loop of the URA3p (β/α)8 barrel and which is distal to both the homodimer interface and catalytic site. After isolation and sequencing of ura3 mutants enriched by the first round of screening, the capacities of each *ura3* variant to convert 5-FOA to 5-fluorouracil, while maintaining activity in the uracil biosynthesis pathway, were performed by integrating each variant into a low-copy vector and transformed this expression cassette into BY4741 Δ*rrm3*. Cells containing either a mutant *ura3* or wild-type *URA3* were then plated on solid media lacking uracil and containing 5-FOA, and relative growth rate was quickly determined by comparing colony size. Strains expressing

ura3(3-5) enabled over 3-fold increases to colony area relative to those expressing the wild-type gene, indicating a decreased propensity of these mutants to convert 5-FOA into 5-fluorouracil while retaining their function in the uracil biosynthesis pathway (Fig 2-8b). Importantly, cells containing this mutant did not exhibit decreased growth rate relative to those containing wild-type *URA3* in uracil-deficient media without 5-FOA, indicating that their increased specificity came with no observable fitness tradeoff under these conditions.

Figure 2-8: Characterization of ICE mutants

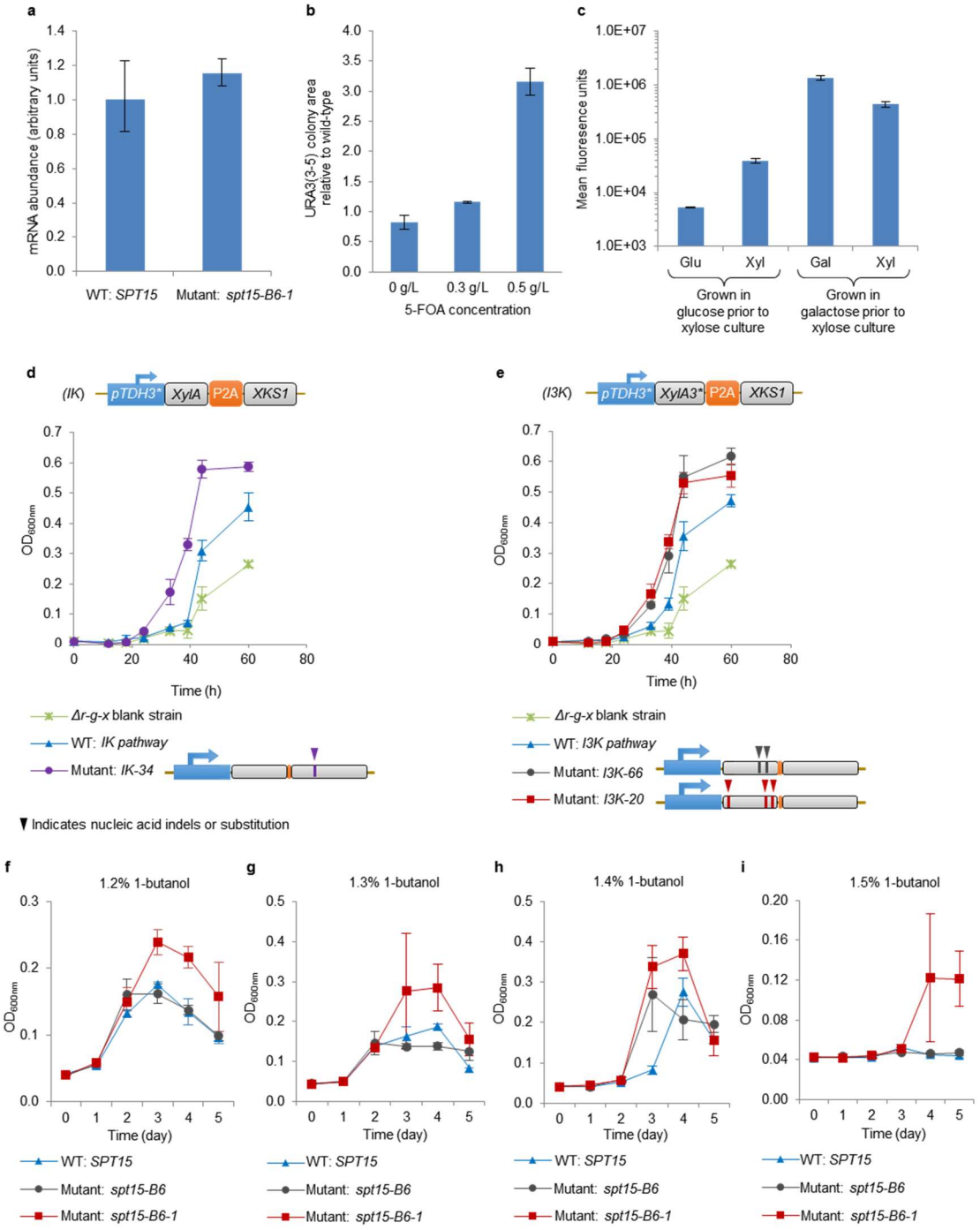
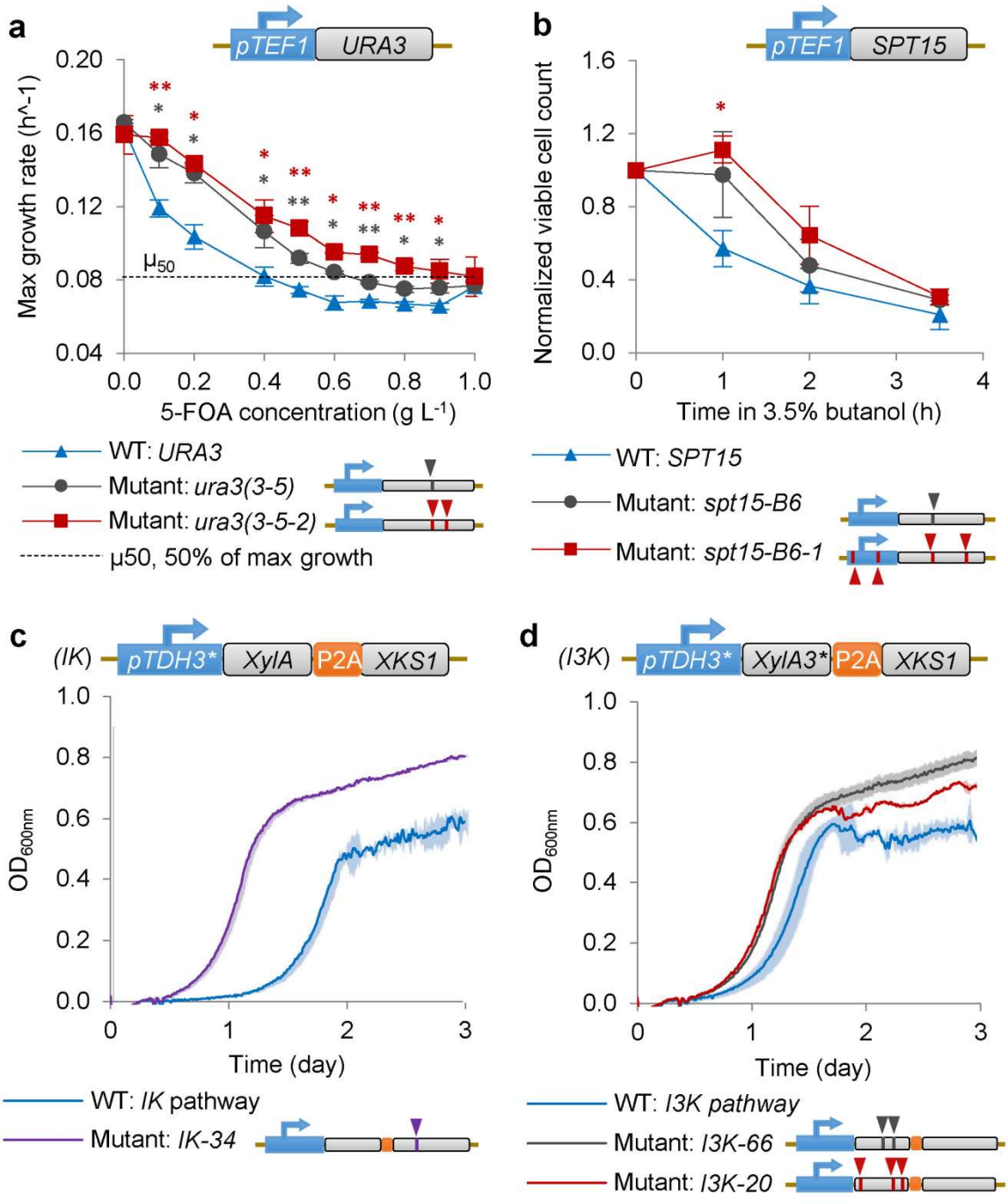


Figure 2-8 (continued): A) Transcriptional rates of *SPT15* mutants. *SPT15* expression levels were measured through qRT-PCR. **B)** Growth of *URA3(3-5)*-containing strains in media lacking uracil and containing 5-FOA. Indicated strains were plated on solid media lacking uracil and containing the indicated concentration of 5-FOA. Colony area was measured using automated image analysis software. Error bars represent standard deviation observed among ~100 analyzed colonies. **C)** Activity of *pGAL1* in xylose-containing media. Strains containing a fluorescent reporter driven by *pGAL1* were cultured in the appropriate medium for several days, then re-seeded in a different carbon source overnight before measurement of fluorescence. Growth test of *xylA* (**D**) and *xylA3** (**E**) pathway mutants. Indicated strains were inoculated into 1 mL of xylose-containing minimal media at a starting optical density of 0.01 and optical density was measured over the course of 3 days. **F)**, Growth test of *spt15* mutants in 1.2% 1-butanol, **G)** 1.3% 1-butanol, **H)** 1.4% 1-butanol, **I)** 1.5% 1-butanol. For (**A**), error bars represent standard deviation of technical triplicates. For (**B-I**), error bars represent the standard deviation of biological triplicates.

ura3(3-5-2), the best-performing mutant from the second round, contained a second coding mutation ($Arg^{186}Lys$) that was likewise distal to the active site and the homodimer interface. *ura3(3-5-2)* was also cloned into a low-copy vector and transformed into BY4741. Cells containing either *ura3(3-5)*, *ura3(3-5-2)*, or wild-type *URA3* were then subjected to growth assays in liquid media lacking uracil but containing from 0 g L^{-1} 5-FOA to 1 g L^{-1} 5-FOA (which is the concentration generally used for efficient counter-selection of the *URA3* marker [107]. Maximum exponential growth rates at each condition were measured in order to compare the relative specificity of each enzyme. In all conditions, strains containing *ura3(3-5-2)* grew the fastest followed by *ura3(3-5)* and finally *URA3*. In particular, cells containing *ura3(3-5)* and *ura3(3-5-2)* attained IC50 values for 5-FOA 1.6- and 2.5-fold higher than strains expressing wild-type *URA3*, respectively (Fig 2-9a). Again, neither mutant strain exhibited decreased growth rate relative to those containing wild-type *URA3* in uracil-deficient media without 5-FOA, confirming this increased specificity did not practically affect native activity.

While further *in vitro* analysis of these variants may reveal the underlying mechanistic basis for improved specificity, this experiment indicated that *in vivo*-generated libraries can be used to evolve substrate specificities using simultaneous mutagenesis and selection.

Figure 2-9: Implementation of *in vivo*-generated libraries for directed evolution of three distinct classes of genetic cargo



▼ indicates location of nucleic acid indels or substitution

Figure 2-9 (continued): Three broad classes of genetic cargo: transcriptional regulators, single enzymes, and multi-gene pathways were used as targets for *in vivo* mutagenesis. **A)** *ura3* variants with altered specificity were characterized by growth rate in uracil-deficient, 5-FOA containing media. **B)** *spt15* variants were characterized by survival in killing concentrations of 1-butanol. **C** and **D)** Xylose pathway variants were characterized through growth rate in xylose-containing media. *IK* pathway variants are pictured in (**C**), and *I3K* pathway variants are pictured in (**D**). In panels (**A**) and (**B**), error bars represent the standard error of the mean derived from biological triplicates, one grey star indicates significantly higher values for the first round mutant versus wild-type at $p<0.05$ via a Mann-Whitney U test, two grey stars indicate significantly higher values for the first round mutant versus wild-type at $p<0.005$ via a Mann-Whitney U test, one red star indicates significantly higher values for the second round mutant versus wild-type at $p<0.05$ via a Mann-Whitney U test, and two red stars indicate significantly higher values for the second round mutant versus wild-type at $p<0.005$ via a Mann-Whitney U test. In panels (**C**) and (**D**), shaded areas represent the standard deviation of biological triplicates.

2.3.3.2 Spt15p Evolution for Improved 1-Butanol Tolerance

Our second target was the gene encoding the global transcriptional regulator Spt15p, the TATA-box binding protein [108]. Traditional *in vitro*-based evolution of this master transcriptional regulator has successfully improved complex phenotypes such as ethanol tolerance [27], but no mutants have been reported which confer increased butanol tolerance. Here, we aimed to use retrotransposon-generated libraries of *spt15* to identify dominant mutants conferring increased tolerance to 1-butanol (see Methods) with each round consisting of mutagenesis in a fresh strain background and selection in non-inducing conditions. Through two subsequent, iterative rounds, we were able to identify a variant (*spt15-B6-1*) that improved the tolerance of a fresh strain nearly 2-Fold after 1 hr in 3.5% 1-butanol (Fig 2-9b, $p<0.05$, Mann-Whitney U test). This mutant also improved growth of a fresh strain of yeast upwards of 44% in 1.4% 1-butanol (Fig 2-8h, $p<0.05$, Mann-Whitney U test). Although we selected for tolerance, not growth rate, the collective performance of this mutant indicated that it significantly outperformed wild-type ($p<0.01$, Fisher's method) in both tolerance and growth characteristics. This mutant contained two coding mutations ($Arg^{98}His$ and $Gly^{192}Ser$) near the DNA-binding domain and two indel mutations in the promoter region (Fig 2-9b). These results indicate the potential of our approach to simultaneously mutate coding and regulatory elements. Moreover, this approach also generated these input libraries in a single, highly automatable step (transfer to galactose-containing media) with significantly reduced labor intensity compared to prior *in vitro* approaches with this transcription factor [27].

The best-performing mutant isolated from the first round, *spt15-B6*, contains a single coding mutation ($Arg^{98}His$) near the DNA-binding domain of this protein, suggesting a putative mode of action. *Spt15-B6-1*, the best-performing mutant from the second round of ICE contained a second coding mutation ($Gly^{192}Ser$) along with two

indels in the TEF1 promoter (Fig 2-9b). This coding sequence mutation, like that of *spt15-B6*, resides in the DNA-binding domain.

After isolation and sequencing of *spt15* mutants enriched by our initial screening, viability analyses of *spt15* mutants were performed by integrating each *spt15* mutant into a low-copy vector and transforming this expression cassette into wild-type BY4741. These strains, along with controls expressing either wild-type SPT15 or an empty vector, were grown to stationary phase and then subjected to a killing concentration of 1-butanol (3.5% by volume). After 0, 1, 2, and 3.5 hours, a small volume of each culture was plated to determine the number of viable cells remaining (see Chapter 8: Materials and Methods). This analysis indicated that cells containing *spt15-B6* exhibited a 1.7-fold higher viability in lethal 1-butanol concentrations compared to wild-type, while cells containing *spt15-B6*-1 exhibited up to 1.95-fold higher viability relative to wild-type under the same conditions (Fig 2-9b).

These mutants were then further tested for any potential growth improvements in butanol-containing media. The same strains, along with controls expressing either wild-type *SPT15* or an empty vector, were grown to stationary phase and then resuspended at a low OD in media containing between 1.2% and 1.5% butanol. These were then grown in anaerobic sealed culture tubes. *Spt15-B6-1* conferred 32% and 44% increased growth over wild-type at 1.3% and 1.4% 1-butanol, respectively ($p<0.05$). As *spt15-B6-1* also contained mutations to *pTEF1*, a qRT-PCR experiment was carried out to investigate potential changes to transcription levels in order to provide insights into the observed phenotype. However, expression measurements indicated no difference between the rate of transcription enabled by *pTEF1* and the promoter contained in *spt15-B6-1* (Fig 2-8a) under these test-tube, exponential growth conditions.

2.3.3.3 Pathway Evolution for Improved Xylose Catabolism

Our third and most complex target was the optimization of a multi-enzyme pathway containing a promoter, an isomerase, and a kinase to enable increased xylose catabolism. Xylose catabolism is an industrially important phenotype for lignocellulose conversion [109] and functional pathways have been established using an evolved xylose isomerase gene from *Piromyces spp*. (*xylA*) [17] and enhanced through overexpression of xylulokinase (*XKS1*) [110]. As either of these two enzymes could serve as rate-limiting steps in this pathway, we established a multi-gene cassette encompassing a strong hybrid *TDH3* promoter (UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-pTDH3, referred to here as *pTDH3**) [111] driving these two coding regions through the use of a P2A site [74] (See Chapter 6). In a second, parallel evolution experiment, we substituted wild-type *XylA* for a previously-identified mutant, *xylA3**, that was shown to confer a 77% increase in xylose consumption rate [17]. These pathway constructs are named IK and I3K, respectively. Although *XylA* has been the focus of prior directed evolution studies, in no case has directed evolution been reported on the entire *XylA-XKS1* pathway simultaneously. For both arrangements, we aimed to identify mutations in a purely in vivo, continuous fashion without intermediate re-transformation steps to demonstrate full continuous evolution. Thus, one continuous round of simultaneous mutagenesis and selection in ICE comprised many potential independent cycles.

After *in vivo* continuous evolution of the IK multi-gene cassette over the course of one week, a superior isolate emerged, IK-34, which contains one coding mutation, $Glu^{164}Lys$, in *xks1*. This isolate displayed a 21% increase in exponential growth rate (Fig 2-9c) over the control as well as an 18-hour shorter lag phase (Fig 2-9c and Fig 2-8d). For I3K, two similarly-performing, superior isolates, I3K-66 and I3K-20, emerged that both contain coding mutations in *xylA3**. In these isolates, *xylA3** contains one

(Ile^{433}Val, in addition to A1029G, a silent nucleotide change) and three (Ala^{48}Ser, Ile^{433}Val, Met^{435}Ile) amino acid substitutions, respectively, and they display 14% and 16% improvements to exponential growth rate, respectively, (Fig 2-9d) concomitantly with a 6-hour decrease in lag time over wild-type I3K. It is interesting to note that the best mutants emerging from these two multi-gene cassettes were distinct (favoring xylose isomerase in one case and xylulokinase in the other) indicating the context-specific nature inherent to directed evolution. We then demonstrated through *in vitro* kinetic assays differences in the function of these mutant enzymes, which indicated isolated mutants indicated increased V_{max} values (0.126 ± 0.008 and 0.134 ± 0.003 μmol min^{-1} mg protein^{-1} for I3K-66 and I3K-20, respectively) compared to wild-type (0.118 ± 0.007 μmol min^{-1} mg protein^{-1}).

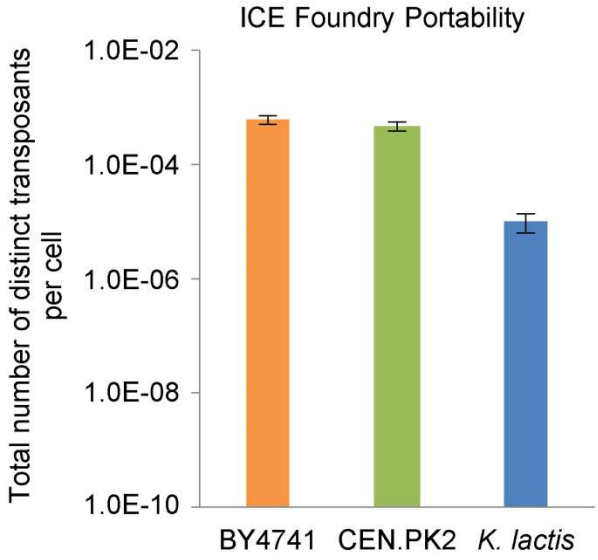
Taken together, performing ICE on entire catabolic pathways resulted in several proficient strains containing multiple mutations which span the entire 4.6 kb pathway after just one week of continuous mutagenesis and selection, which is currently the longest pathway for which in vivo continuous directed evolution has been undertaken. As a result, we realized a significant acceleration to the process of creating pathway-wide mutants as compared to classical directed evolution techniques that require creating separate libraries of each gene and reassembling pathways *in vivo* [71].

2.3.4 Implementation of ICE in Alternative Yeast Strains

Finally, we wished to investigate the portability of the retrotransposon-assisted mutagenesis approach across divergent species of yeast. Specifically, we observed that alternative strains of *S. cerevisiae* such as CEN.PK2, as well as divergent species of yeast such as *Kluyveromyces lactis* were also able to support replication of the synthetic

retroelement with only minimal modifications (Fig 2-10). However, beneficial knockouts found in *S. cerevisiae* BY4741 did not transfer to CEN.PK2, and the lack of a knockout collection in this strain prohibited us from exploring this organism for further improved backgrounds. As Ty1 activity in *K. lactis* has only been recently demonstrated [112] and the precise mechanism of Ty1 transposition in *K. lactis* is still uncharacterized, these preliminary results motivate future work investigating the species-dependence of Ty1. Nevertheless, this result indicates that the general approach of utilizing retrotransposons to undertake in vivo mutagenesis may be expanded to any other eukaryote which supports LTR retrotransposon activity [92], making this a potentially powerful, broad-host approach.

Figure 2-10: Transposition rates for alternate yeast strains



Transposition rates for wild type BY4741 (low OD plasmid induction), CEN.PK2 (low OD plasmid induction), and *K. lactis* (high OD genomic induction). Error bars at the low OD condition represent 95% confidence intervals obtained from biological triplicates, while error bars at the high OD condition represent the standard deviation of biological triplicates.

2.4 Discussion and Conclusions

In this work, we developed an *in vivo* mutagenesis system for yeast and applied it to the directed evolution of small-molecule-converting enzymes, regulatory proteins, and metabolic pathways. Significantly, we have shown that this enables large, diverse mutant libraries in a continuous process which is significantly faster, more effective, cheaper, and less labor-intensive than traditional in vitro techniques. Moreover, we have demonstrated that the ICE approach can surpass a traditional error-prone PCR library in both library and clone performance. Motivated by its utility demonstrated by this work, several challenges of ICE in its current form can be addressed to further its speed and effectiveness. Specifically, increasing the error rate for small cargo, increasing library sizes when evolving more than 5 kb, reducing the reintegration of mutants into alternative loci, and reducing the potential for concurrent strain adaptation under certain modes of selection are goals for future designs. Nevertheless, this approach in its current instantiation is capable of supporting continuous evolution of parts and pathways as demonstrated here. This ICE approach complements powerful continuous and genome-scale engineering techniques in other organisms [67, 87, 90], can interface with any screening/selection technique, and is the first demonstration of an in vivo continuous evolution approach for optimization of small-molecule-converting enzymes and pathways (of which *XylA-XKS1* is the longest yet reported for any *in vivo* continuous directed evolution approach). Taken together, this work introduces the retroelement-assisted continuous evolution paradigm, demonstrates its utility for the directed evolution of a wide variety of phenotypes, and indicates its unique potential to enable powerful new applications for the rapid evolution of cellular components across varied eukaryotic hosts.

Chapter 3: Further Development and Characterization of *in vivo* Continuous Evolution

3.1 Chapter Summary

The utility and range of *in vivo* Continuous Evolution was validated by rapidly evolving several novel phenotypes. Several limitations remain, however, especially the relatively low rates of retrotransposition and errors on a per-cell basis. Additional experiments focused on improving this system by increasing the potential library size, either by increasing the rate of retrotransposition or by increasing the error rate of the cargo DNA. In addition, novel elements with potential to improve the system were incorporated and tested. Primary efforts involved the reverse transcriptase, either through protein engineering or by replacing it wholesale with the well-characterized and error-prone HIV reverse transcriptase. Furthermore, concurrent overexpression of other factors, including native genes known to influence retrotransposition and *trans*-acting Ty1 elements, was investigated. Finally, the effects of various parts of the ICE system previously validated, such as the artificial intron and the galactose induction conditions, were more fully characterized. While no significant improvements over the original ICE system were found in these experiments, they served to more fully characterize how ICE operates and identify several unique opportunities for further investigations.

3.2 Introduction

Collectively, the results of the previous chapter indicate that *in vivo* Continuous Evolution (ICE) is a powerful and easy-to-implement method for undertaking directed evolution in yeast. Significantly, large and diverse mutant libraries are generated in only 3 days, using a process significantly faster and less labor-intensive than traditional *in vitro* approaches. Additionally, ICE seamlessly interfaces with any downstream screening

and isolation methods and is thus applicable to any screenable phenotype. This approach complements high-throughput engineering techniques in other organisms and is the first demonstration of a continuous evolution approach for improvement of catalytic enzymes.

However, several targets for improvement were also identified. First, ICE throughput seems to be currently limited by the reverse transcriptase activity and fidelity, as a relatively low rate of retrotransposition and mutation rate results in the vast majority of cells in a library culture *not* containing a mutant variant. If the Ty1RT enzyme could be engineered for higher processivity or decreased fidelity, it could potentially increase the library size obtainable. It may also be possible that the overall performance of ICE may be improved by replacing it with a more error-prone reverse transcriptase such as that of HIV [96]. It may also be possible to incorporate exogenous elements to better control the ICE cycle, such as Cas9 or T7 RNA polymerase.

In addition, there are several other factors to be investigated that may identify potential bottlenecks and areas for improvement. Such factors, including the effects of the artificial intron, the culturing conditions during ICE induction, and the termination of the ICE mRNA produced during initial transcription, need to be further studied beyond the initial validation. Exploring these effects more fully can reveal simple ways to improve the utility of an ICE approach.

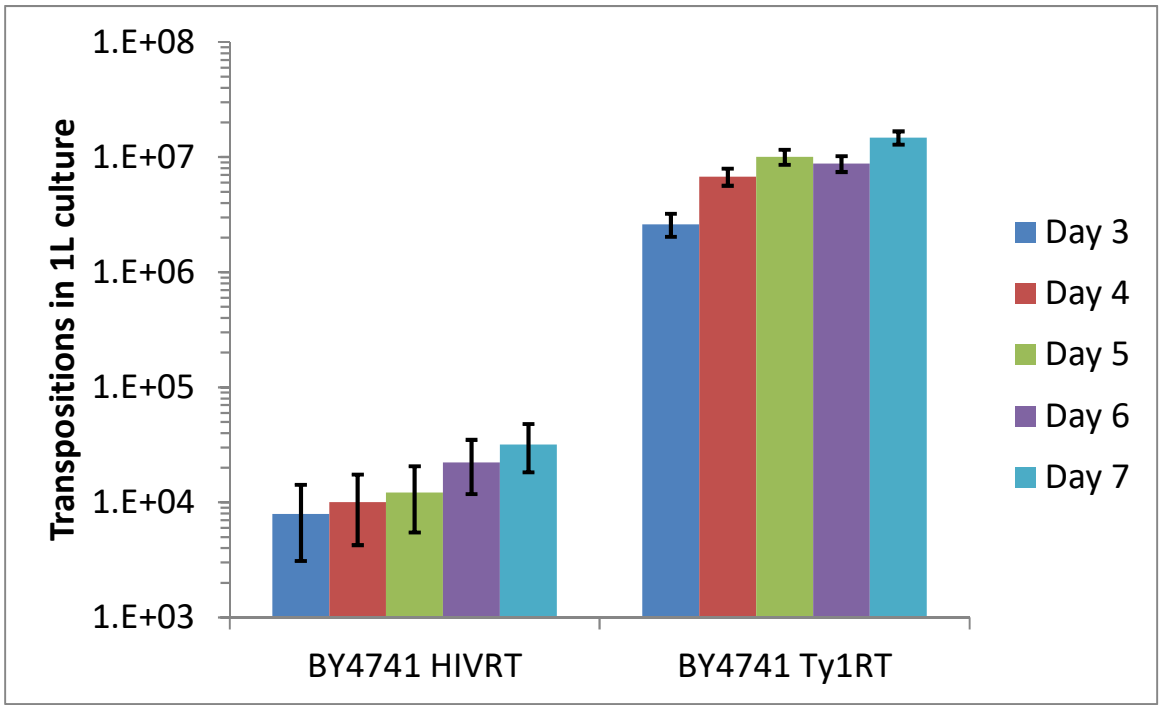
3.3 RESULTS

3.3.1 Incorporating HIV Reverse Transcriptase into ICE

The first and most extensive effort to more fully control ICE mutagenesis was the incorporation of the HIV reverse transcriptase in place of the native Ty1 enzyme. Prior reports have suggested that the Ty1 reverse transcriptase is modular, and that it is

possible to simply replace the native enzyme with the HIV version [113]. Since the HIV reverse transcriptase is well-characterized and known to be significantly more error-prone than the native Ty1 enzyme [96], it was hypothesized that imported HIV into the ICE system could increase the potential library size achievable by enabling a higher error-rate during reverse transcription. However, during initial tests, this HIV-ICE construct exhibited significantly lower activity than the ICE construct based on the native enzymes (Fig 3-1).

Figure 3-1: Transposition rate with different reverse transcriptase

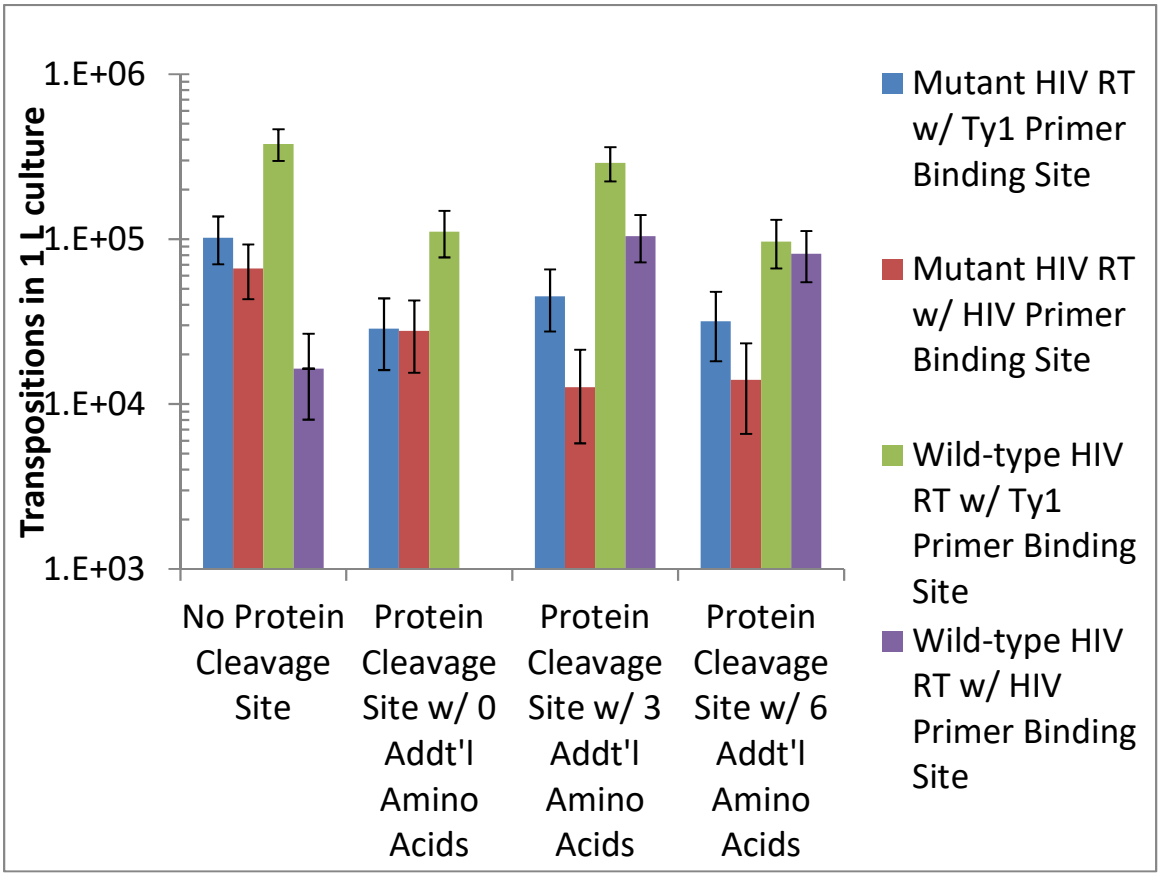


ICE constructs were made using the HIV reverse transcriptase (left) or the native Ty1 reverse transcriptase (right). Each were induced in galactose-containing media for 3-7 days, and plated on uracil-dropout media to test transposition rate. Those strains incorporating the HIV reverse transcriptase exhibited significantly reduced levels of retrotransposition

One possible explanation for this lack of activity was hypothesized to be inefficient polypeptide cleavage of the integrase/reverse transcriptase system. While this cleavage is required for both HIV and Ty1 systems, the cleavage site is different, and Ty1 is not known to act efficiently on the HIV sequence [114].

In an effort to create a retroelement system incorporating HIV RT with a higher retrotransposition activity, several variants of the HIV-ICE construct were made. First, two variants of HIV RT were made, both a wild-type and one containing mutations known to influence fidelity [115]. Next, the retroelement RT primer binding sites were changed from those native to Ty1 into those native to HIV. Finally, the sequence between the integrase gene and the reverse transcriptase gene, coding for the protease cleavage site wherein the polypeptide is cleaved to form the two mature enzymes, was altered to code for the native Ty1 cleavage site with either 0, 3, or 6 additional amino acids from the Ty1 RT gene downstream. It was hypothesized that if the HIV RT was inactive due to improper protease cleavage, including more native Ty1 RT amino acids at this site would facilitate more effective cleavage and thus more active enzyme. Sixteen constructs were made through combinations of each of these factors; each was then transformed into a By4741 strain and the transposition rate under galactose induction was measured.

Figure 3-2: Transposition rate with various HIV Reverse Transcriptase variants

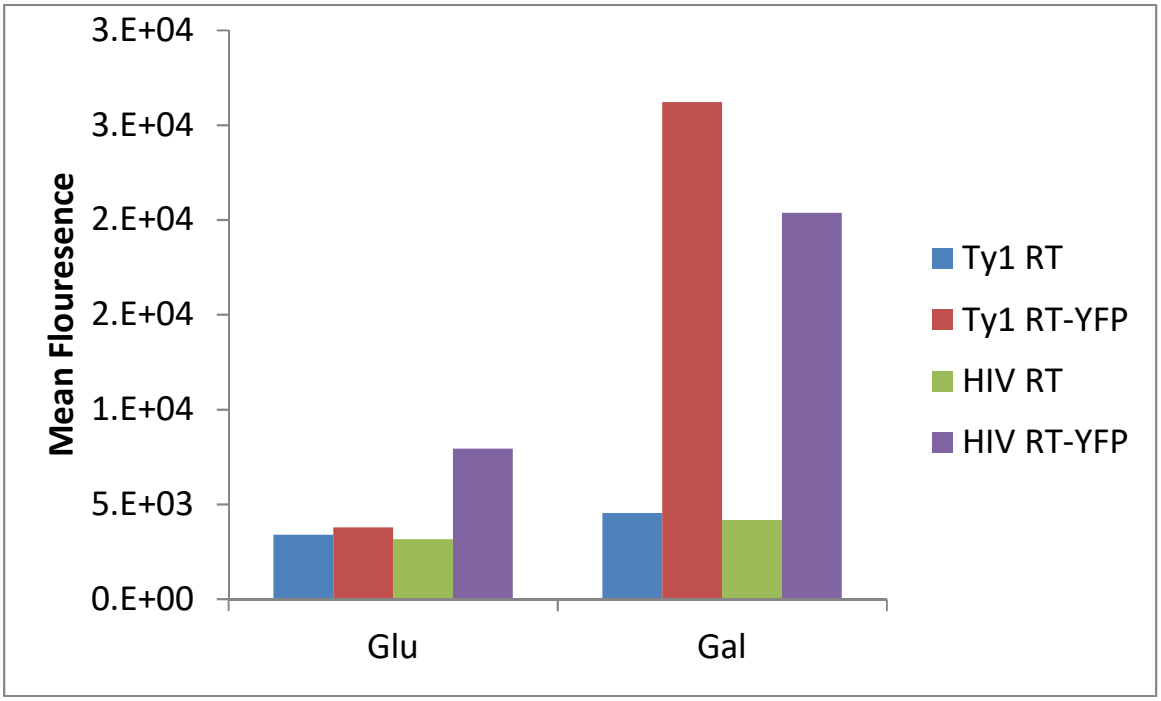


Mutant and WT HIV RT was cloned into ICE with several variations: different primer binding sites and different protein cleavage site. There was no clear trend with a significantly higher rate of retrotransposition

No clear pattern was found in the retrotransposition activity of these HIV constructs (Fig 3-2). There was some variation between strains, but none displayed a significantly elevated retrotransposition rate. These data seem to indicate that the low activity previously observed with HIV-ICE is probably not due to improper protease cleavage [113], but more likely due to HIV RT interfering with reverse transcription catalyzed by natively present Ty1 RT. The variable activity observed in the different variants may be only due to the extent to which each mutant interferes with this process.

To ensure this is not due to a simple expression issue, and in order to quantify and characterize the expression of both Ty1 and HIV reverse transcriptase as part of the ICE system, fluorescent fusion proteins were made by adding the YFP gene downstream of the reverse transcriptase gene in the optimized retroelement, with a linker region between them to allow proper folding and fluorescence. These constructs were transformed into BY4741, and fluorescence was measured by flow cytometry after growth in either glucose or galactose. The results (Fig 3-3) indicate that both Ty1 and HIV RT are being expressed only on galactose induction, as expected. Unfortunately, this implies that the inability to utilize a HIV RT-containing retroelement is not due solely to inadequate expression, but that the activity of HIV RT is simply suboptimal in the yeast cell environment.

Figure 3-3: Fluorescence of RT-YFP Fusion Proteins



Strains expressing either the normal ICE cassette or ICE cassettes containing YFP-linked RT fusion proteins, grown in either glucose-containing media (left) or galactose-containing media (right), and measured by flow cytometry. These results suggest that both enzymes are being expressed under galactose exposure, as expected.

3.3.2 Ty1 Reverse Transcriptase Mutagenesis

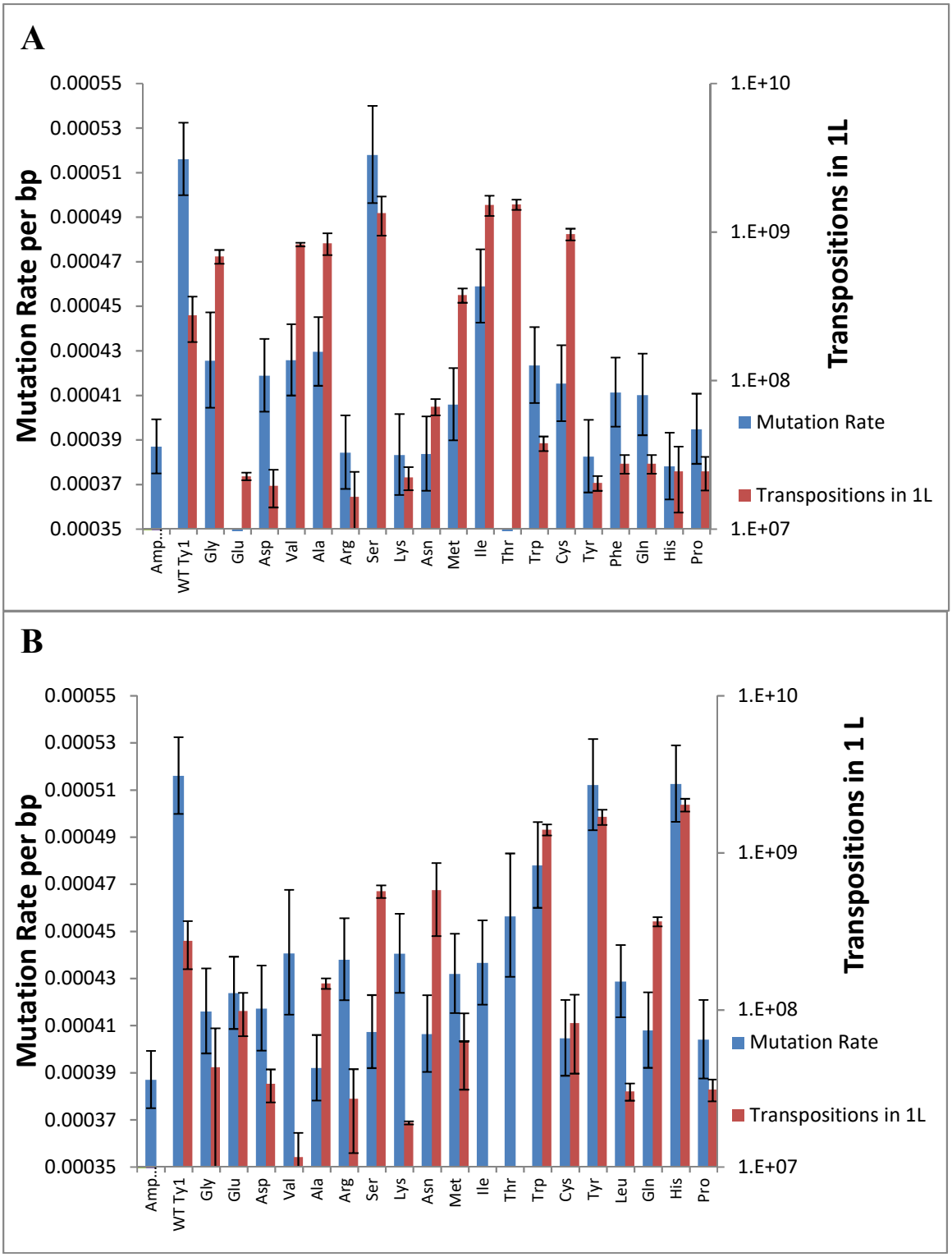
With the difficulties in adapting HIV RT into the ICE cassette, another approach focused on altering the native Ty1 system instead. One potential strategy is to engineer the Ty1 reverse transcriptase to increase the error rate of the while maintaining its relatively high transposition rate.

The first approach focused on creating a library of enzyme variants by performing site-specific saturation mutagenesis on several residues. The amino acids in positions 145, 225, and 226 are highly conserved across different reverse transcriptase enzymes and have been identified as playing a key role in fidelity in the homologous region of

HIV [116-120]. A full saturation mutagenesis library was created for each of these three sites, consisting of 19 variants containing each amino acid substitution. These 57 plasmids were then transformed independently into the BY4741 *ΔRRM3* strain, and the transposition rate of each was measured. In addition to assaying transposition rate, mutational analysis was conducted by extracting total DNA after galactose induction, with *URA3* sequences resulting from transposition events amplified via PCR. PCR products were then pooled and next-generation sequencing was performed on the mixture. This was done as in the original next-generation sequencing experiments that quantified the ICE mutagenesis rate [121](See Chapter 2).

It was found that while several mutations increased or maintained transposition rate, many reduced activity significantly (Fig 3-4). However, an increase in both transposition rate and mutation rate was observed in strains with three different point mutations in the Ty1 RT: L145S, F225R and F225H. It is estimated that a system incorporating the Ty1 RT with any of these three mutations could potentially increase the library of unique transposants by approximately 5-fold.

Figure 3-4: Transposition rate and mutation rate of Ty1 RT mutants



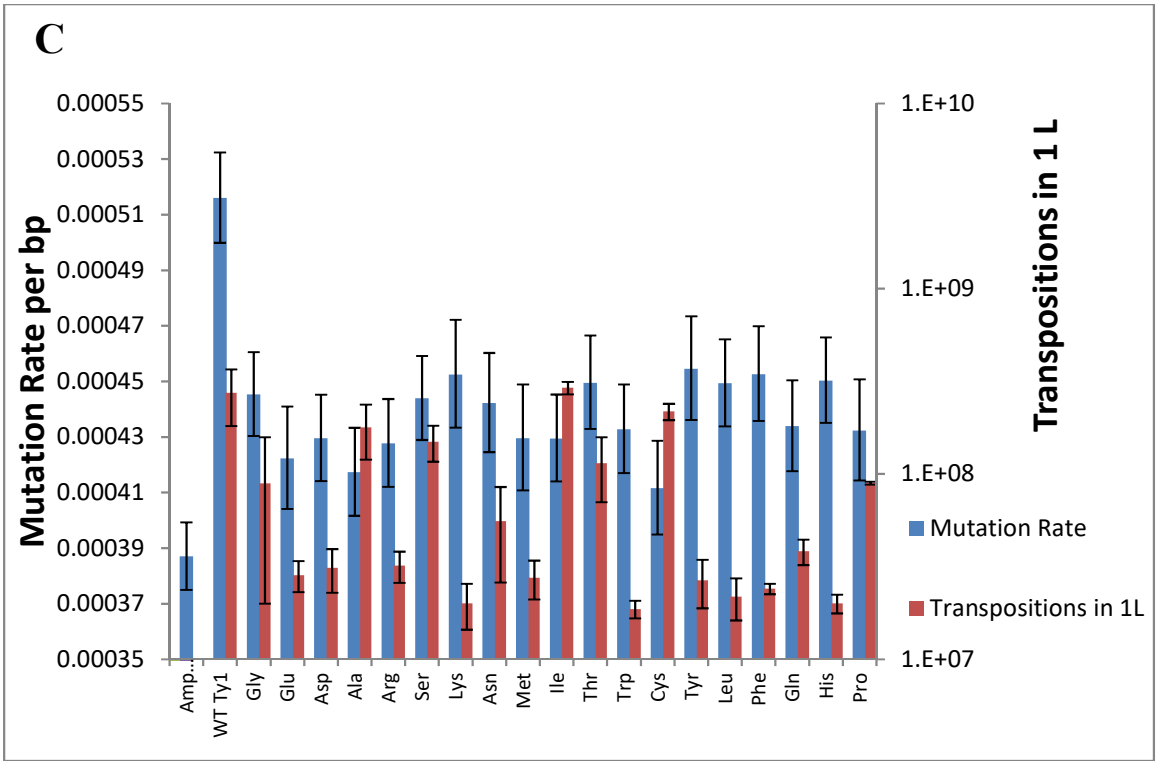


Figure 3-4 (continued): The mutation rate (blue, left axis) as assayed by next-generation sequencing and the transposition rate (red, right axis) is displayed for each amino acid substitution of **A)** Leu145, **B)** Phe225, and **C)** Val226.

Next, these three mutations were transferred to the *dKanMX* system to simultaneously measure *in vivo* activity and mutagenesis of these systems. As discussed in Chapter 2, *dKanmx* is a non-functional *KanMX* antibiotic marker constructed by the insertion of an artificial stop codon. This is placed as the ICE cargo with each independent Ty1 RT variant. After induction, the culture is plated on G418-containing media, allowing growth only if the point mutation is reverted. This serves to measure both activity and error-rate of the Ty1 RT variants, since the number of colonies will be directly related to each.

At the same time, several additional Ty1 RT variants were created, including K93R, R94K, and Y151A. These mutations were chosen by selecting additional residues

that have been shown to influence error-ate in HIV RT, and aligning the sequences of the two RTs to find homology. For example, the equivalent mutation of Y151A in HIV (Y115A) was shown to increase errors ~4-fold [122]. Similarly, the lysine and arginine residues at 93 and 94 were both shown to play a role in fidelity [115, 123]. Lastly, two double-mutants were made by combining the three previously studied mutations described above: L145S and F225R, as well as L145S and F225H.

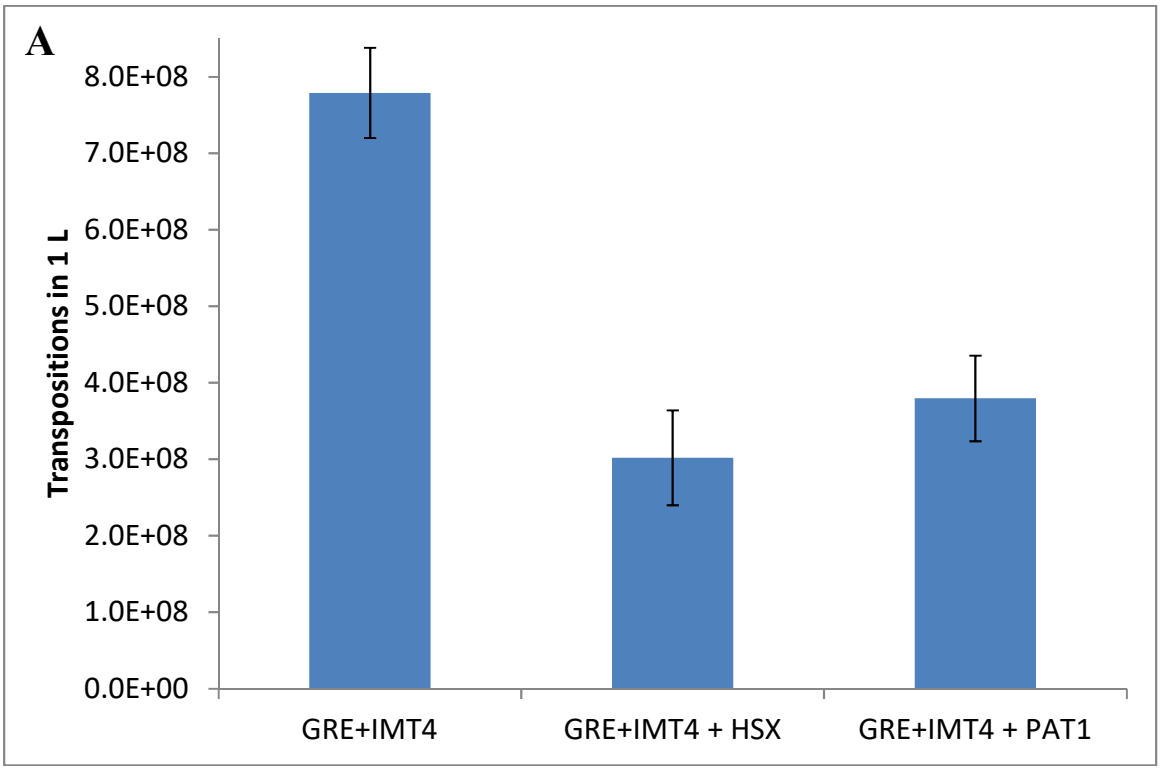
These eight Ty1 RT variants were inserted into the ICE cassette containing *dKanMX* cargo, and RT activity and fidelity was measured by inducing with galactose followed by subsequent plating on G418-containing media. Unfortunately, none of these variants resulted in a higher number of reversion colonies, suggesting that none exhibited improved characteristics. It is possible that these mutations may have increased activity and simultaneously reduced the error-rate, or increased the error-rate but decreased activity; in either case, the number of reversions would not necessarily increase. This experiment suggests that the wild-type already results in the largest possible library size. Until more study of the Ty1 RT is completed, perhaps including the protein structure being elucidated, it is difficult to do any further rational engineering of this enzyme.

3.3.3 Gene Overexpressions for Improvement of Transposition Rate

Another potential path explored was the improvement Ty1 activity through strain engineering. Previously, a large collection of gene knockouts linked to Ty1 suppression were investigated (See Chapter 2). We also explored overexpressing endogenous Ty1 activating or supporting genes, in the hopes of relieving a bottleneck or supporting higher rates of activity. Several genes known to be associated with higher Ty1 activity were chosen, including *hsx1*, *pat1*, *dhh1*, *kem1*, and *lsm1* [124]. Each of these genes was

independently amplified and cloned to a high-copy plasmid with a strong GPD promoter[3], and then transformed into the strain that exhibited the highest rate of ICE activity: By4741 *ΔRRM3* with tRNA overexpression plasmid and an integrated copy of the ICE retroelement. These strains were then induced to test transposition rate. In each case, the concurrent overexpression had reduced or similar activity compared to the control with no concurrent overexpression (Fig 3-5). These results suggest that none of these targets improved the transposition rate and thus library size of ICE.

Figure 3-5: Transposition rate with concurrent overexpressions



[3] Note: *hsx1* encodes an arginine tRNA, so the native PolIII promoter was used instead of GPD

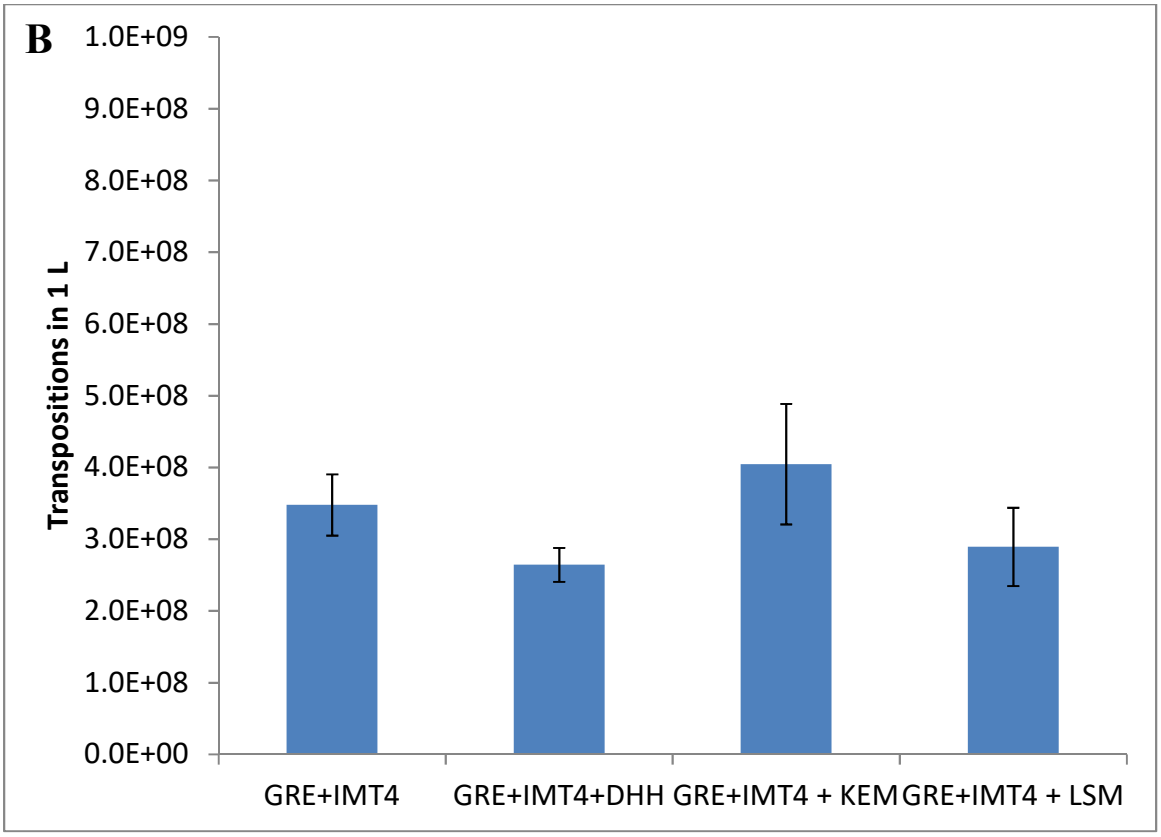


Figure 3-5 (continued): "GRE" denotes an integrated copy of ICE with intron-interrupted *pTEF-URA3*cargo, where and IMT4 denotes the high-copy IMT4 plasmid. each additional overexpression is included as a second plasmid to the base strain (left-most); **A)** displays the overexpression of *hsx1* and *pat1* while **B)** shows *dhh1, kem1*, and *lsm1*. Note that these experiments were carried out on different days with different induction times, which may explain the difference in the control strain results.

3.3.4 Integrase Engineering for Improvement of Transposition Rate

It has been previously observed that during ICE induction, cDNA generated may be integrated into the genome at loci other than the original ICE cassette integration. Because such genomic integration may generate effects which are not related to the function of the enzyme being mutated, it is desirable to minimize the extent of this phenomenon. However, it is known that in Ty1 the integrase and reverse transcriptase are processed as a polyprotein which is subsequently cleaved [125]. In an attempt to

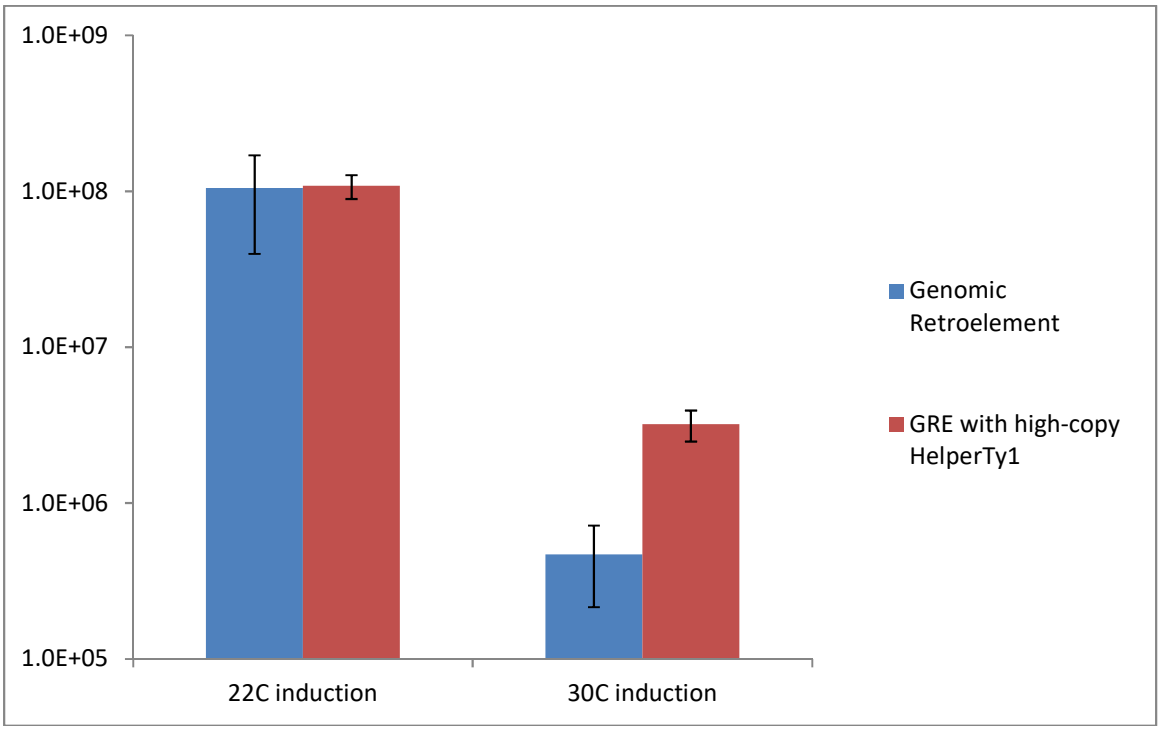
reduce genomic integration without compromising reverse transcription activity, a Ty1 integrase mutation previously found to inhibit integrase activity was constructed [126]. The construct was then retransformed into a BY4741 strain and the transposition rate test was carried out. It was found that total transposition activity was significantly reduced when this integrase mutation was present, rendering this mutant unsuitable for the ICE system.

3.3.5 Developing a Disassociated Reverse Transcriptase System

Another effort to improve the ICE transposition rate utilizes a dissociated reverse transcriptase system, with *trans*-acting Ty1 elements that facilitate more efficient retrotransposition. One previously reported construct, called "Helper"-Ty1, consisting of the 5' portion of the native Ty1 retroelement, including the reverse transcriptase but excluding the 3' long terminal repeat. It has previously been shown that expression of this construct can catalyze the transposition of distal retroelements containing both long terminal repeats but not the reverse transcriptase [127]. It was hypothesized that overexpressing this element on a plasmid could help catalyze a higher rate of retrotransposition of the genomic element, by allowing multiple copies of the reverse transcriptase to be expressed without having to include multiple copies of the cargo. The element was placed on a vector and co-expressed in cells containing a genomic integration of the retroelement with the URA3-intron reporter gene, and then these cells were induced at high-OD stationary phase to measure retrotransposition. It was determined that although this co-expression did improve transposition at 30ºC, it did not have this effect at 22ºC (Fig 3-6). It is hypothesized that although overexpression of this

element may improve retrotransposition when proteolytic processing is inhibited at 30ºC, this is not the rate limiting step at the more stable 22ºC (See Chapter 2).

Figure 3-6: Transposition of genomically-integrated ICE retroelement with co-expression of "Helper"-Ty1



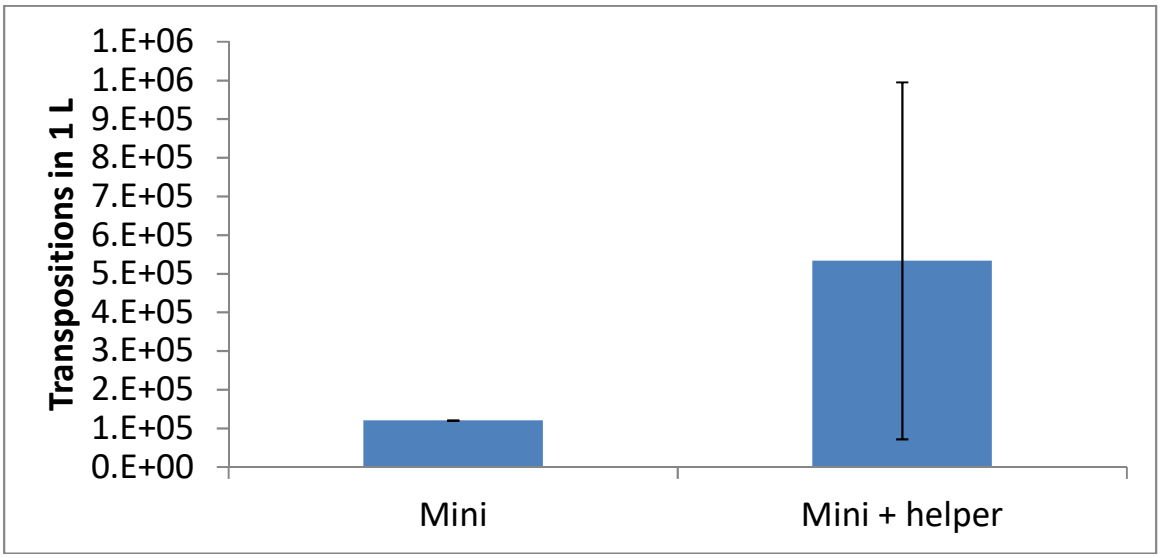
Helper-Ty1 is co-overexpressed on a high-copy plasmid; the control strain has no plasmid.

It was next hypothesized that this Helper element could instead be used to catalyze the retrotransposition of a synthetic retroelement that did not contain the reverse transcriptase or other native proteins, but consisting solely of the cargo to be evolved and the LTR. Such a system was also shown to be functional, with the short synthetic element denoted a "Mini"-Ty1 [127]. The advantages of such a system would be several: the retrotransposition rate could conceivably be higher, since the length of the RNA

transcript to be reverse transcribed would be much shorter; the reverse transcriptase gene itself would not be reverse transcribed, avoiding any deleterious mutations from accumulating that could inhibit a second round of retrotransposition; and finally such a system could potentially avoid the re-integration elsewhere in the genome, instead favoring recombination with the original shortened element.

To test this idea, a "mini" retroelement was constructed from the original ICE system by removing all native genes upstream of the TEF-URA3 insert, leaving both long terminal repeats. Next, this synthetic mini-retroelement was integrated in the genome, and a high-copy "Helper" construct was co-transformed. The strain containing only the mini-retroelement and a strain with the helper construct co-transformed were both assayed with high-OD galactose induction. Unfortunately, both strains had an extremely low rate of retrotransposition, near to the background levels previously observed (Fig 3-7); this indicates the Helper element was probably not significantly catalyzing retrotransposition. This result suggests a disassociated reverse transcriptase system is not yet ready for use in ICE evolution.

Figure 3-7: Transposition of genomically-integrated "mini-"ICE retroelement with co-expression of "Helper"-Ty1



3.3.6 Utilizing Orthogonal T7 RNA Polymerase to Improve Error Rate

Another strategy investigated was to incorporate the use of T7 RNA polymerase to transcribe the mRNA of the ICE cassette, instead of using the native PolII inducible promoter *pGAL1* to directly transcribe the retroelement. The rationale for this idea relied on the highly-characterized nature of T7, including the description of particularly error-prone variants [128]. If errors could be introduced during transcription as well as during reverse transcription, the overall mutation rate of a single ICE cycle could potentially be significantly expanded. Since error-prone T7 RNAP variants have been reported to exhibit errors at a rate ~10-fold higher than one ICE cycle [128], successfully incorporating it would vastly improve the achievable library size. However, it is not clear if mRNA transcribed by T7 would be successfully exported from the nucleus and localized to the virus-like particles in which Ty1 reverse transcription is catalyzed. Since it has been shown that Ty1 mRNA transcribed in *trans* can be successfully reverse

transcribed [127], it was hypothesized that T7-transcribed mRNA could also be reverse transcribed by endogenous Ty1 elements, even if the T7-transcibed mRNA could not be translated.

To test this, the original ICE system was modified to incorporate wild-type T7 RNAP [129]; the *pGAL1* promoter driving expression of the ICE retroelement was swapped out for the T7 promoter, and then T7 RNAP was expressed on a high-copy plasmid under the control of *pGAL1*, such that ICE cycles could still be induced by growth in galactose culture. Then, transposition was tested by induction as before. Interestingly, there were *no* prototrophic colonies observed, suggesting activity was completely eliminated. While it was expected that transposition rate would be reduced, it seems that simply incorporating T7 RNAP as-is does not allow for any retrotransposition events at all. While it is understood that T7-transcribed mRNA cannot be translated in yeast, this experiment suggests that this mRNA cannot even be recognized by the native endogenous Ty1 machinery present either. Further engineering must be considered if T7 would be used in the ICE system.

3.3.7 Utilizing CRISPR-Cas9 to Direct cDNA re-Integration

As discussed above, it has been observed that the cDNA generated during a cycle of ICE can re-integrate into another location in the genome. In order to improve the efficiency of the first round of retrotransposition and more efficiently re-direct cDNA integration back to the original locus, Cas9 was expressed in conjunction with a gRNA specific for the artificial intron of URA3. Through this strategy, it was hoped that the mutated cDNA generated during reverse transcription would site-specifically integrate to its original location (through the double-stranded break generated by Cas9) via

homologous recombination. To this end, mutagenesis and Cas9 expression were induced simultaneously to yeast encoding a genomic copy of the retroelements through growth on galactose. If Cas9 functioned as desired, each viable cell surviving after induction would contain an intact *URA3* gene. However, we observed a large excess of cells which contained no intact *URA3* gene after induction, resulting in a ~10-fold lower transposition rate relative to cells which did not express Cas9. However, upon analysis of the URA^+ colonies, it was observed that isolated transposants had replaced the original copy of *URA3* in the retroelements with the intronless copy in every colony tested, which is in contrast to the situation for strains which do not express Cas9 which seem to mainly integrate the intronless copy of URA3 elsewhere in the genome. Thus, although Cas9 expression can improve the specificity of cDNA re-integration, this specificity comes at the cost of a reduced transposition rate and library size. Because of this tradeoff, it was determined that Cas9 expression would not be utilized unless re-integration specificity became a major limiting factor for ICE in the future.

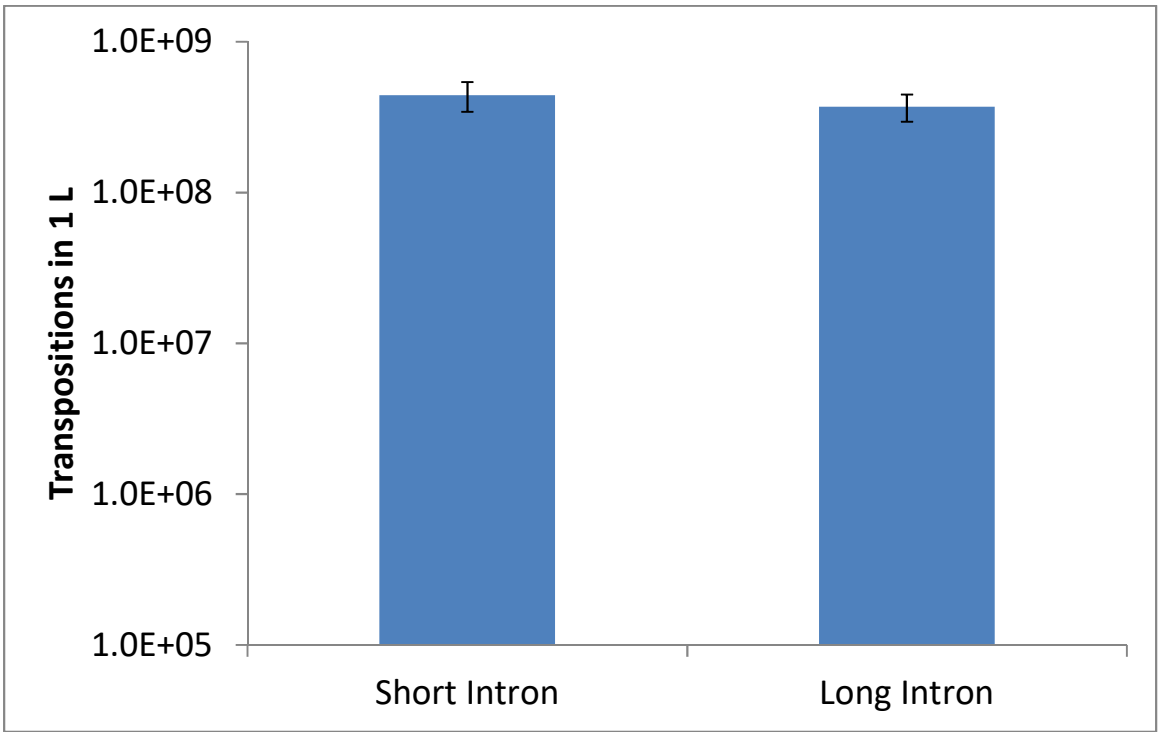
3.3.8 Characterizing the Effect of the Artificial Intron on Transposition Rate

The original construct used for assaying rate of retrotransposition utilizes an artificial intron in the reverse orientation to allow quantification of transposition events, and this artificial intron was designed with multiple restriction sites to allow further manipulation. However, in the intron cassettes used for evolution, a slightly different artificial intron was utilized, incorporating only one restriction site, making it slightly shorter – 73 bp vs 104 bp. Furthermore, the artificial intron in evolution cassettes (see Chapter 2) interrupts the start codon of the gene being evolved, while the cassette used

for retrotransposition quantification has the intron located in the middle of the *URA3* gene.

To quantify the effects, if any, the artificial intron length and placement has on the rate of retrotransposition, several variants of the reporter system were constructed. First, the current artificial intron was simply replaced with the shorter version used in evolution cassettes. This construct was then transformed into the By4741 Δ*RRM3* strain, and retrotransposition was assayed with the galactose induction. This test confirmed that both retroelement constructs exhibited the same rate of retrotransposition (Fig 3-8).

Figure 3-8: Transposition rate of ICE system with different intron interrupting *URA3* reporter



The short intron contains a single added restriction site, and is 73 bp; the long intron has several, and is 104 bp long. These were tested in the same conditions to demonstrate that the intron length does not materially affect transposition.

Next, two more constructs were made that incorporated both of the two intron versions into the *URA3* start codon, in order to test if this effects the rate of retrotransposition. These were also transformed into the By4741 *ΔRRM3* strain, and retrotransposition was assayed with galactose induction. Interestingly, those strains with the start codon interrupted were *all* able to grow on uracil dropout medium, making the retrotransposition rate impossible to quantify. It is possible that interrupting the start codon with an intron still allows the expression of a slightly shorter open reading frame, resulting in at least partial activity even without retrotransposition and intron excision. This is an important lesson for the construction of ICE retroelements; it is possible that if an artificial intron is placed too close to the start of a gene, a partially-active version may still be able to be expressed without retrotransposition. While this is almost certainly gene-dependent, future constructs should have artificial introns inserted into the middle of the gene to ensure only retrotransposed copies are able to be expressed.

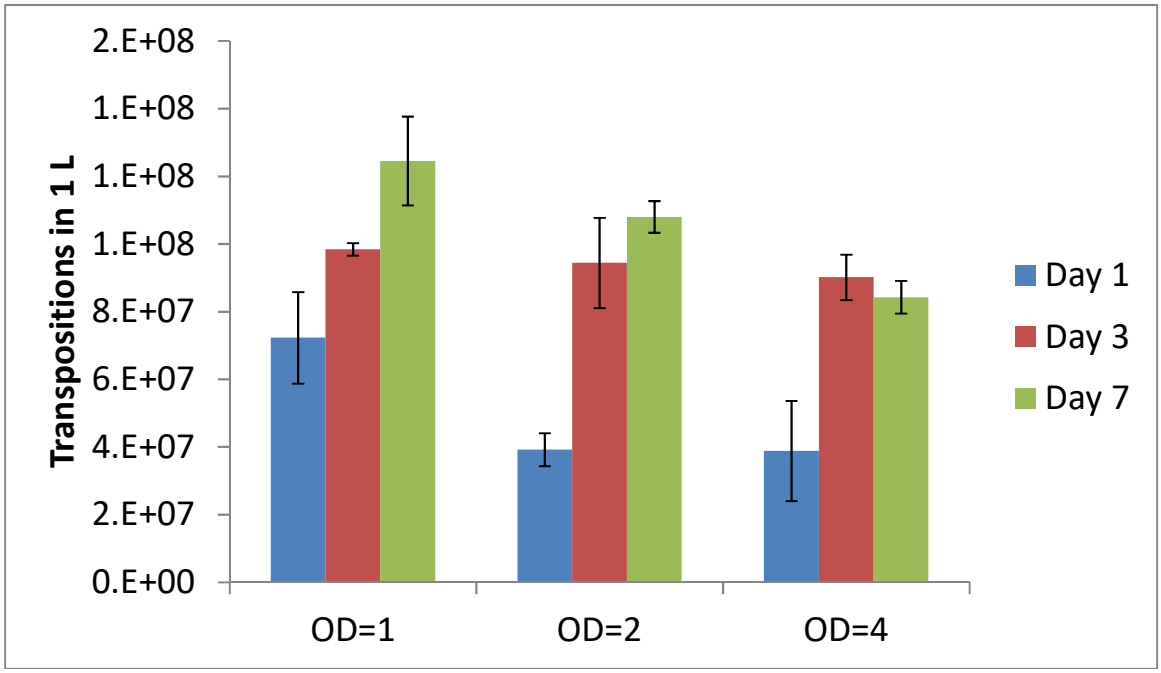
3.3.9 Further Characterization of Induction Culturing Conditions and the Effects on Transposition Rate

Previously, it was established that culturing conditions during galactose induction can have significant effects on the overall rate of transposition, with cultures induced at high-OD exhibiting retrotransposition rates orders of magnitude higher than those cultured at low-OD (see Chapter 2). Several additional culturing conditions were also investigated, to determine if it would be possible to manipulate conditions to foster further improvement in this activity.

First, the starting culture OD was simply increased, resulting in more concentrated conditions during galactose induction. Cultures were suspended in

galactose-containing media at OD=1.0, 2.0 or 4.0, then induced for 1, 3, or 7 days before plating. Those cultures induced at higher OD did not exhibit any increase in transposition events, with the maximum number of events observed in those cells at OD=1.0, the condition used for all evolutionary experiments (Fig 3-9). Interestingly, because there were more cells present in the more densely cultured cells, this indicates that the transposition rate on a per-cell basis actually *decreased* at these conditions. These results suggest that induction at OD of approximately 1.0 is apparently the optimum condition for maximizing retrotransposition.

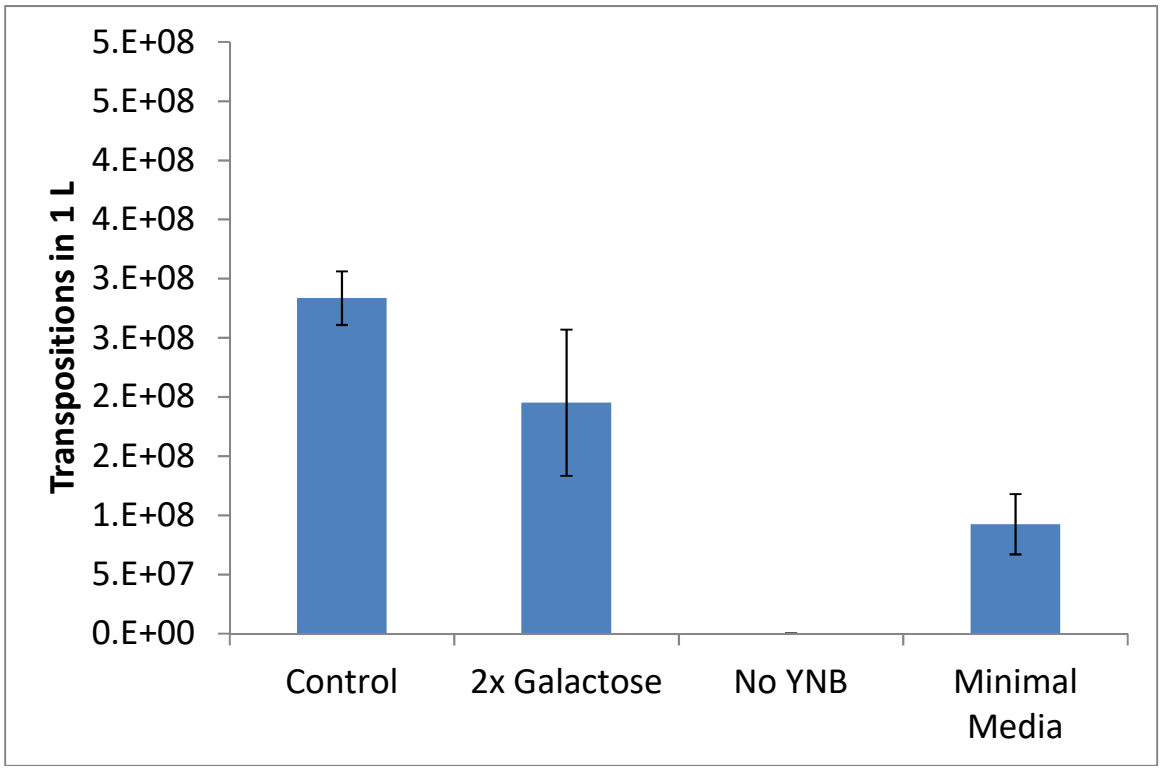
Figure 3-9: Transposition rate at with induction at different cell densities



Strains with the integrated ICE cassette containing *pTEF-URA3*-intron as cargo were induced in galactose at various initial ODs. Cells were plated on uracil-dropout media after 1, 3, and 7 days of induction, and the transposition rate was quantified by counting colonies. Interestingly, higher cell density did not lead to more retrotransposition events.

Next, a similar experiment was undertaken by altering the media makeup of the induction culture. Several variations were tested: using media with twice the concentration of galactose (40g/L), media with no YNB, and "minimal" media, containing only galactose, auxotrophic amino acids, and YNB. In each case, the rate of transposition was reduced from that of the control (Fig 3-10); in fact, removing YNB eliminated transposition entirely. Again, these results suggest that induction with standard CSM media may maximize retrotransposition.

Figure 3-10: Transposition rates with different induction media composition



Strains with the integrated ICE cassette containing *pTEF-URA3*-intron as cargo were induced in galactose with either a more concentrated carbon source ("2x galactose", which is 40g/L), no YNB, or a minimal media containing only galactose, auxotrophic amino acids, and YNB. None of these conditions improved the overall rate of transposition.

3.3.10 Effects of mRNA Termination on ICE Transposition

Another possibility investigated was the potential to increase the rate of retrotransposition by increasing levels of mRNA with high-efficiency terminators. It has been shown that using different terminators in synthetic biology applications can significantly affect activity, potentially by increasing mRNA half-life [130]. It was theorized that if ICE mRNA half-life was increased, it could potentially result in a higher overall rate of transposition, by allowing more opportunities for reverse transcription prior to RNA degradation. To this end, several high-efficiency native terminators were inserted into the ICE cassette immediately following the 3' LTR, including CPS, IDP, and PRM [130]. These were transformed into yeast and the transposition rate measured as described previously. However, none of these strains exhibited any change in overall rate of transposition when compared to the control construct, which contains no explicit terminator. These results suggest that the 3' LTR may act efficiently as a terminator for the initial mRNA transcription and adding additional termination sequences does not impact the process.

3.4 Discussion and Conclusions

In this work, extensive characterization and investigation of various factors with the potential to improve the process of *in vivo* Continuous Evolution (ICE) was carried out. The original ICE system proved a powerful method of creating new libraries in yeast, but it could be substantially improved if there were methods to increase the overall potential library size. Here, we investigated many potential factors that might be part of such a method. While there was no obviously significant improvement, the additional

characterization of these factors further elucidates the limitations and process of ICE, and informs future directions of research.

For example, key experiments carried out here investigated simple changes that had potential to significantly impact the retrotransposition process, such as the artificial intron and the induction conditions. Here it is shown that neither of these factors limit the rate of retrotransposition, establishing that the process used in the original experiments were close to optimal in this respect.

Furthermore, multiple concurrent overexpression targets were investigated, including native genes that promote retrotransposition and *trans*-acting Ty1 elements that can enable more efficient retrotransposition of the ICE element. Such overexpression targets would seem intuitively to offer the most likely path for increasing the rate of retrotransposition, and thus the potential library size enabled by ICE; however, we show that none of these have a significant effect. This suggests that either the retrotransposition process is not rate-limiting or that simply overexpressing supporting genes concurrently does not have a significant effect over simple galactose-induced expression.

Finally, a substantial amount of effort was spent exploring the possibility of engineering the Ty1 reverse transcriptase to increase its error rate. This was done mostly by sing protein homology to the well-studied HIV reverse transcriptase, since the Ty1 enzyme has no reported crystal structure. There is some substantial homology, however, especially in conserved regions known to affect fidelity. Unfortunately, none of these promising targets enabled higher rates of transposition or error-rates, as measured by the *dKanMX* reversion assay. It is certainly possible that several such mutants *did* reduce fidelity, but did so at a cost to processivity of the reverse transcription reaction; such a tradeoff would actually reduce the library size. Without further fundamental

understanding of the Ty1 system, including more study of the reverse transcriptase enzyme, there does not seem to be an opportunity for further engineering.

These experiments also reveal how the ICE process may be improved. While it seems difficult to further optimize the complex retrotransposition process, there remains promise in incorporating orthogonal layers of control, such as Cas9. By targeting Cas9 to the ICE cargo, we were able to better control cDNA reintegration, which otherwise did not seem possible. Future experiments will focus on experimentation with such orthogonal systems, including T7 RNA polymerase, which could potentially expand the capabilities of ICE significantly.

Chapter 4: Development of Spinach-based Metabolite Biosensors

4.1 Chapter Summary

Directed evolution techniques are still fundamentally limited by screening capabilities, especially for the discovery of non-growth production phenotypes. While extremely large libraries of variants can be created, traditional screening techniques for metabolite production are low-throughput. Protein-based biosensors help bridge this gap but can be difficult and slow to engineer. Here, we develop a novel generalizable method for creating Spinach-based RNA biosensors. Spinach is a short RNA aptamer that binds to a non-fluorescent dye, creating a fluorescent complex; by combining secondary aptamers into its structure it has been shown that this fluorescence can be made dependent on the concentration of the secondary aptamer's target. By expanding this capability using a modeling approach in collaboration with Dr. Howard Salis, we demonstrate that such a system can be adapted to rapidly create functional *in vitro* biosensors for a wide range of molecules. The approach is validated by applying it to two aptamers never before functionalized in a Spinach setting, tyrosine and 2,4-dinitrotoluene. We further validate the modeling approach by creating a range of tyrosine sensors, demonstrating that their functionality matches that predicted by the model. This approach enables the rapid creation of novel biosensors for a huge variety of targets that will significantly improve directed evolution efforts.

4.2 Introduction

Directed evolution for pathway metabolic engineering requires an extremely large library of mutations and an equivalently high-throughput screening method in order to identify desired variants. While growth-based selection is often a powerful tool, it can

often be limited, especially when the desired goal is to evolve production phenotypes that actually inhibit growth. Traditional techniques such as GC, HPLC, and MS can provide excellent data allowing for the selection of mutants improving product generation; however, they are low-throughput and costly when compared to *in vitro* assays.

As an alternative to growth-based selection, biosensors have been developed to link metabolite production to some reporter, often fluorescence [131, 132]. However, there remains a lack of generalizable biosensor generation techniques that can be quickly and easily adapted to a different or novel target. Generally, biosensors must be independently developed for each new target, a process that can be lengthy, difficult, costly, and not guaranteed of success. When engineering an entire pathway through evolutionary techniques, there may exist the need to screen multiple metabolite intermediates; developing novel biosensors for each undoubtedly limits the process.

To address this, we utilize RNA aptamer biosensors with the potential to be used for high-throughput screening of large libraries. In particular, RNA aptamers linked to "Spinach", a previously constructed RNA molecule capable of inducible fluorescence, can be used as sensors of metabolite concentration [133]. Such a sensor links the *in vitro* concentration of the metabolite of interest to fluorescence, which can be rapidly and easily assayed. Critically, such a system is generalizable, in that such a sensor can be rapidly and easily created for any molecule for which an aptamer has been reported. Since there have been hundreds of such aptamers described for a variety of molecules [134], the ability to use Spinach-based biosensors is not limited to any one specific molecule or even class of molecules.

Previously, however, creating Spinach-linked sensors has not been straightforward. Their function relies on conditional folding: when the metabolite is bound, Spinach will properly fold and fluoresce, but when the metabolite is not bound

Spinach will not fold correctly. In prior efforts, aptamer sequences that resulted in this conditional folding were created through a process of trial-and-error, by manually examining predicted folding behavior, creating multiple sequences that could plausibly function as desired, and then synthesizing each and testing them, selecting those that worked at highest efficiency [133].

Here, we improve on the development process for Spinach-based metabolite biosensors through a modeling approach, in conjunction with Dr. Howard Salis at Penn State Univeristy, in which the optimal sequence for coupling a metabolite-binding aptamer to Spinach can be identified computationally. Importantly, this approach can also help tune both the sensitivity of such a sensor as well as its brightness, by accurately modeling ΔG values of each bound/unbound permutation. This process results in a powerful tool for rapidly creating novel biosensors necessary for evolutionary engineering of metabolic pathways.

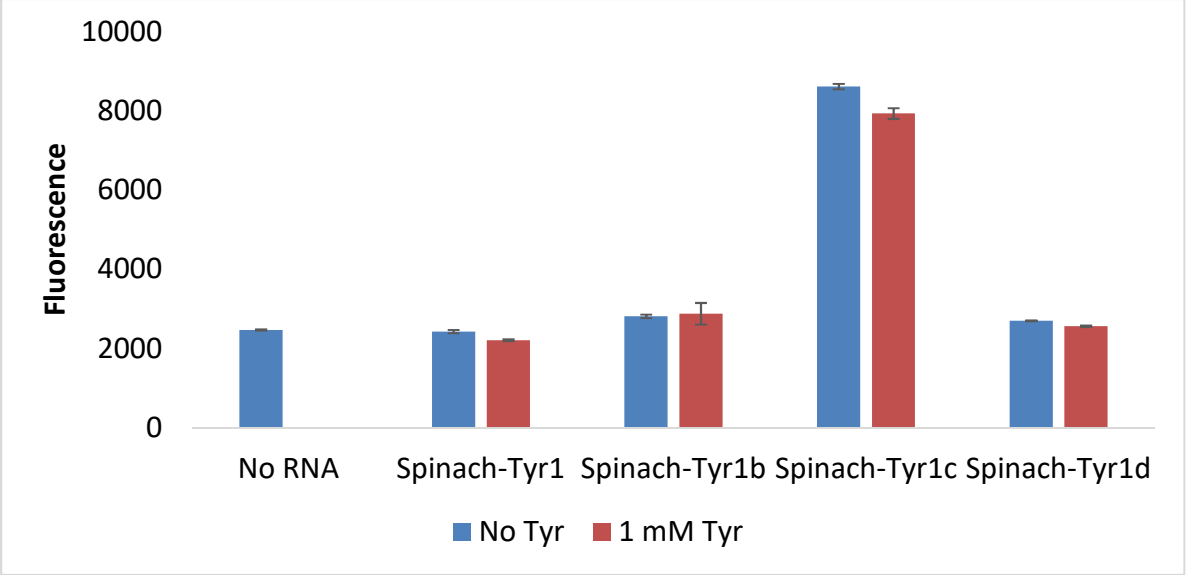
4.3 RESULTS

4.3.1 Initial Construction of Spinach-Tyrosine Biosensor

As a proof-of-concept demonstration of Spinach aptamers' screening capability, tyrosine was chosen as a production target. Tyrosine (like other aromatic amino acids) is subject to significant, tight regulatory control in yeast and thus the pathway suffers from extremely poor yields and titers. Molecules derived from the shikimate pathway (including tyrosine) have multiple uses as polymer precursors, nutritional supplements, and therapeutic agents [135]. Importantly, a previously reported Tyrosine aptamer "Tyr1" binds relatively tightly to tyrosine, with a K_d of 23uM, a physiologically relevant concentration [136].

The first Spinach-Tyrosine sensor was created by simply inserting the Tyr1 aptamer sequence into the Spinach stem loop as demonstrated previously. Fluorescence of this Spinach-Tyr1 aptamer was minimal, however (Fig 4-1), so several variants were constructed by altering the "linker" region to disfavor secondary structure. To do this, the predicted folding structure of various sequence perturbations were examined using mfold [137], again as was reported previously. Three variants were created in this process, with one variant (Tyr1c) significantly brighter than the original aptamer. However, none of the three had any sensitivity to tyrosine presence, indicating that the Spinach structure was stable irrespective of the Tyr1 aptamer (Fig 4-1).

Figure 4-1: Fluorescence of first series of Spinach-Tyr1 aptamers tested

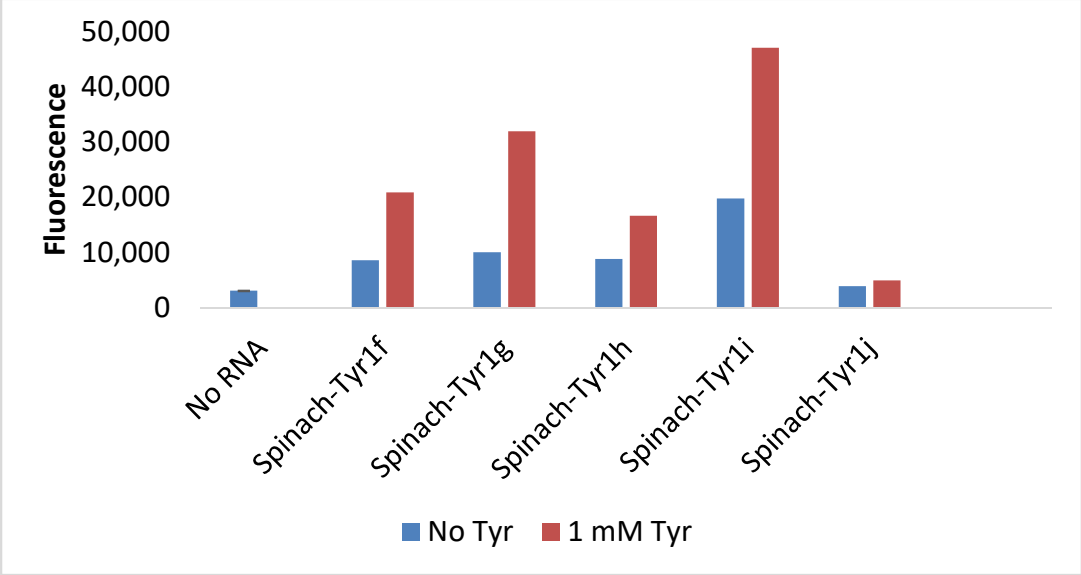


Blue bars represent samples with RNA, DFHBI, and buffer, but no tyrosine added. Red bars represent the same conditions with 1 mM tyrosine present.

Tyr1c was then used as a scaffold for a second round of variants, again altering the linker region in base pairs predicted to contribute to secondary structure. Five

second-round variants were made, and several exhibited tyrosine sensitivity, with several exhibiting both a high differentiation in signal as well as brightness comparable to Spinach aptamers reported (Fig 4-2).

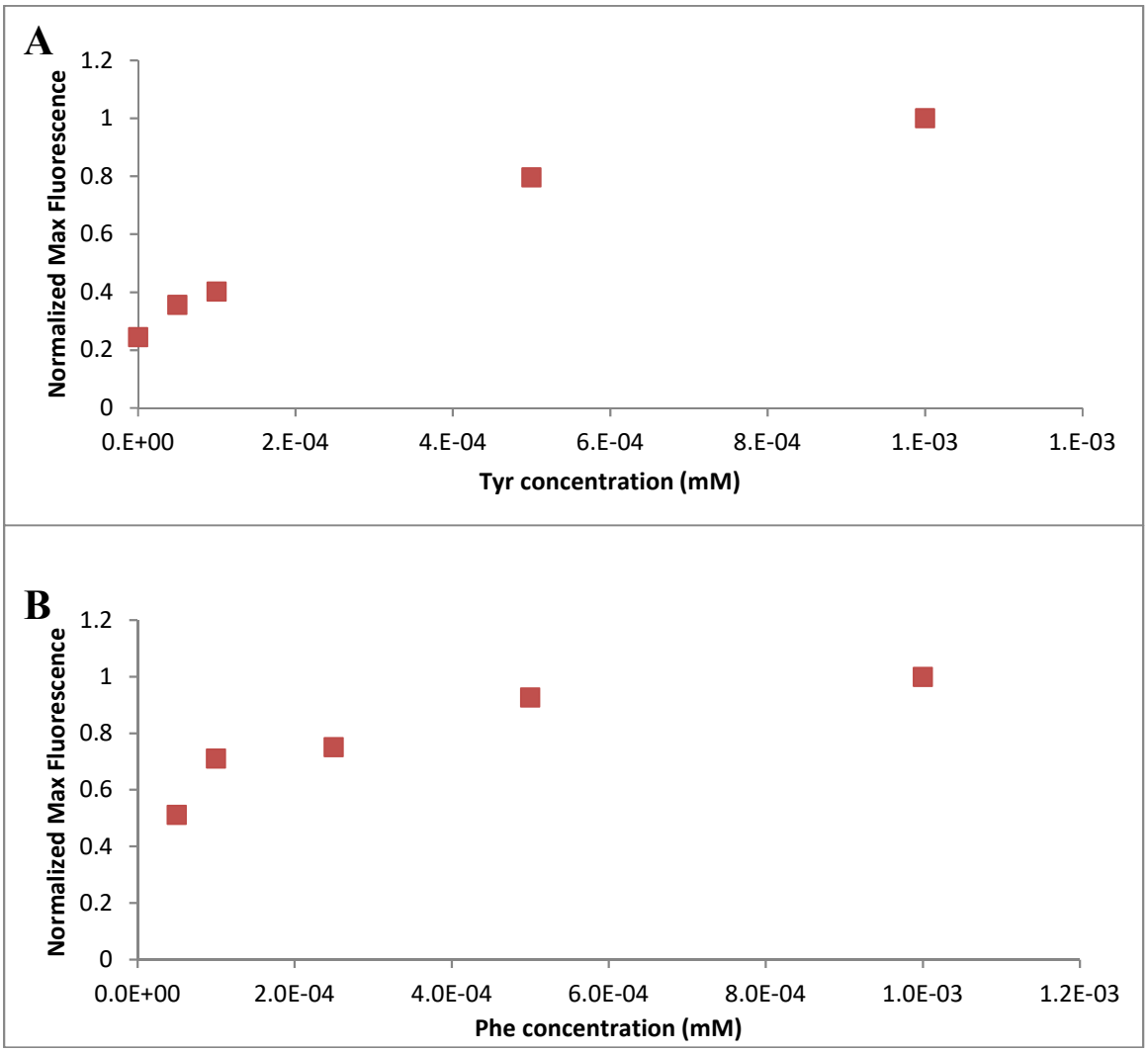
Figure 4-2: Fluorescence of second series of Spinach-Tyr1 aptamers tested



Blue bars represent samples with RNA, DFHBI, and buffer, but no tyrosine added. Red bars represent the same conditions with 1 mM tyrosine present.

Next, Tyr1i was further characterized. It was shown to be sensitive to tyrosine concentrations as low as 0.05 mM, with a signal strongly correlated with concentration (Fig 4-3a). The same aptamer could also detect other aromatic amino acids phenylalanine and tryptophan as well, again in relation to concentration (Fig 4-3b-c). There was no signal in the presence of non-aromatic control leucine, though, confirming its suitability as a biosensor for aromatic amino acids.

Figure 4-3: Fluorescence of Spinach-Tyr1i with varying concentration of aromatic amino acid present



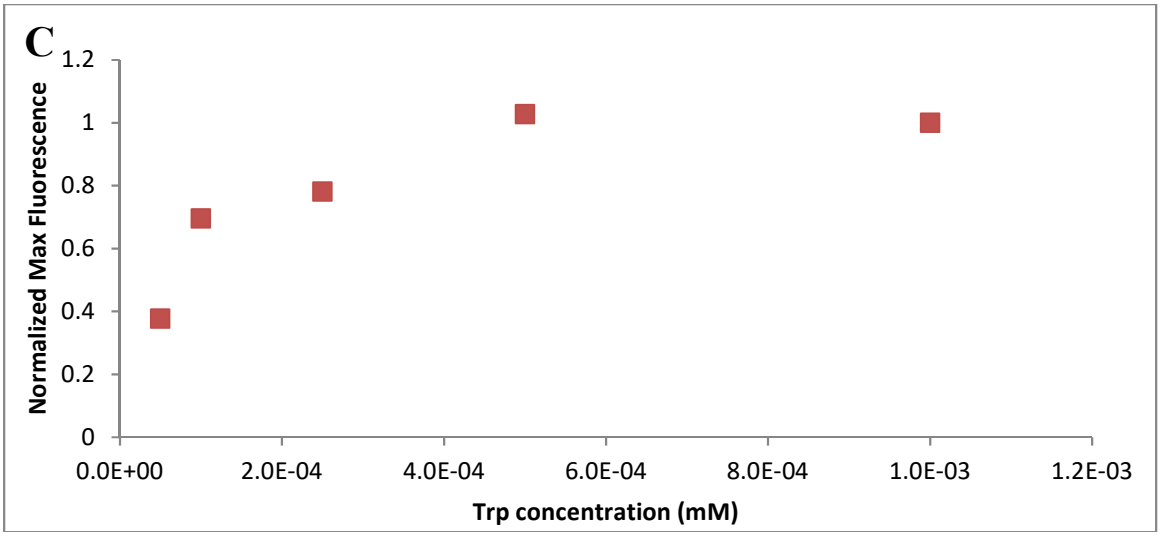
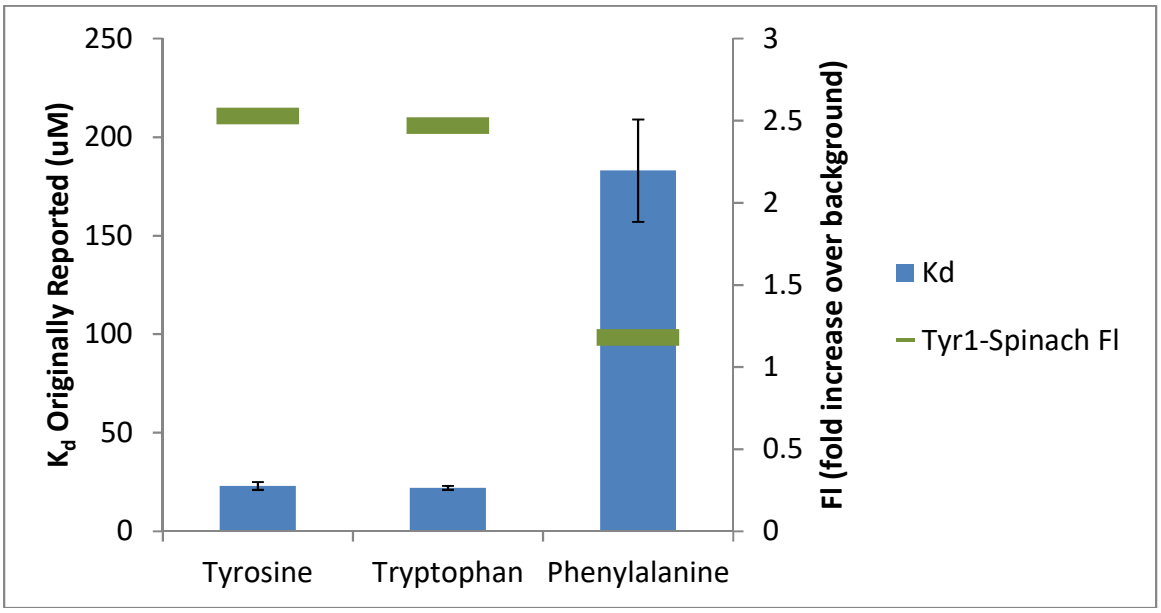


Figure 4-3 (continued): Values are normalized to fluorescence observed with the metabolite at 1mM. **A)** tyrosine, **B)** phenylalanine, and **C)** tryptophan

Critically, these differing sensitivities to the aromatic amino acids matches the original binding affinities measured for the aptamer Tyr1 (Fig 4-4). This important finding confirms that incorporating metabolite-binding aptamers into a Spinach-based biosensor does not significantly affect the binding properties. This is essential to the generalizability of such a system; while many aptamers have been characterized, the binding properties must remain consistent when bound to Spinach if they are to be useful as biosensors. The results with the Tyr1 system suggest this is the case.

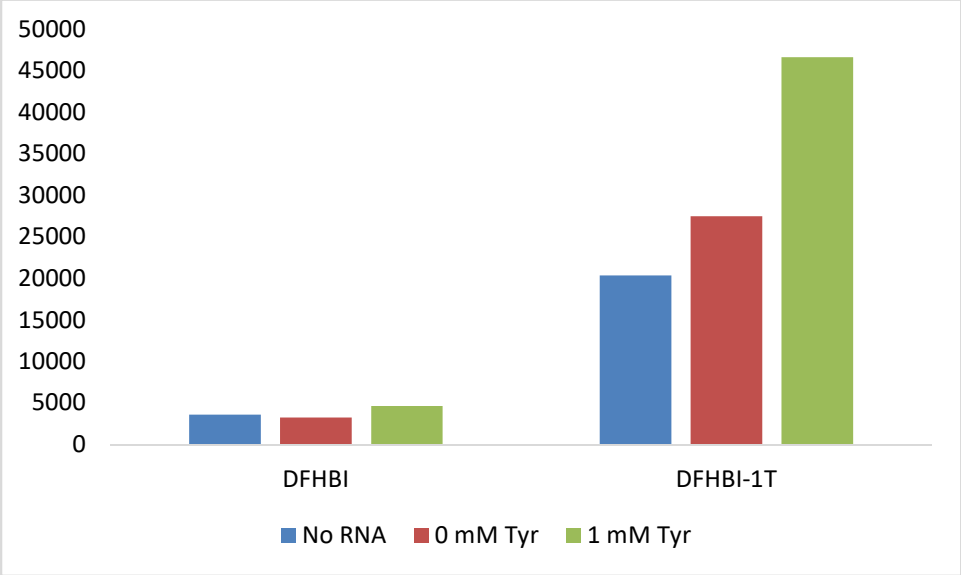
Figure 4-4: Comparison of Tyr1 K_d values to Spinach-Tyr1 sensitivities for three aromatic amino acids



The Tyr1 aptamer was reported to be much more sensitive to tyrosine and tryptophan compared to phenylalanine. When incorporated into Spinach, the resulting fluorescence observed matched this observation: a much stronger signal was observed with tyrosine and tryptophan when compared to the same concentration of phenylalanine (1mM for each).

Finally, these aptamers were also tested using the then-newly available version of the DFHBI dye used to detect Spinach aptamers, called DFHBI-1T, which had been reported to have a brighter signal [138]. Importantly, the absorbance/emission spectra of this system that more closely matched that of GFP, which is very useful in a biosensor, so that existing optics system for high-throughput analysis could be utilized. The Spinach-Tyr1 was adapted to this system by incorporating the relevant point mutations to the Spinach sequence, creating Spinach2-Tyr1. When tested, this version of aptamer resulted in both a stronger signal and better signal differentiation between negative and positive samples than that using the original dye (Fig 4-5). After this was confirmed, further experiments solely utilized Spinach2 with DFHBI-1T.

Figure 4-5: Tyr1i tested with either the original Spinach sequence and the dye DFHBI or the updated sequence Spinach2 and the dye DFHBI-1T



Each experiment used the same concentration of RNA and the respective dye. Fluorescence was read at the peak excitation/emission wavelengths of each dye: 460/500nm for DFHBI and 482/510 for DFHBI-1T. While the background fluorescence of DFHBI-1T is much higher, the signal and the fold-increase over background is also much higher.

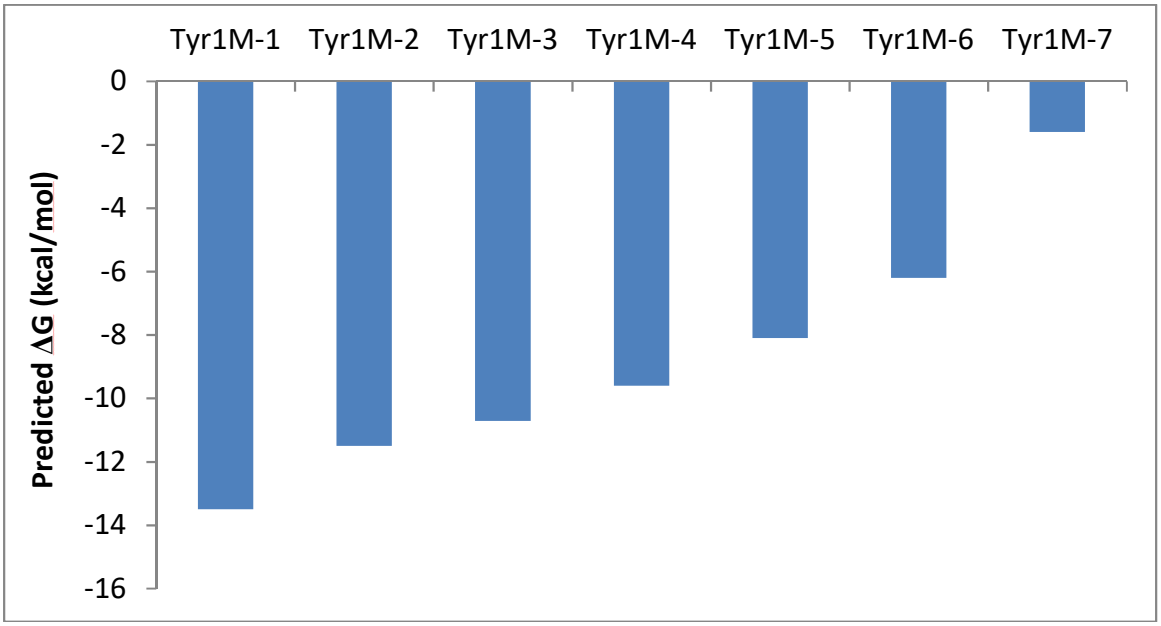
4.3.2 Modeling Linker Regions of Spinach-Tyrosine Biosensor

In order to improve the process of creating novel metabolite-spinach biosensors, a modeling approach was undertaken in collaboration with Dr. Howard Salis at Penn State Univeristy. In this approach, a range of possible "linker" sequences was tested computationally to calculate ΔG of binding in each permutation, including the aptamer to the dye, the aptamer to the metabolite, and the aptamer to both dye and metabolite. By comparing the values for thousands of potential linker sequences, an "optimal" series of sequences can be determined by finding linker regions that maximized the change of ΔG between binding dye and binding dye and metabolite, called ΔΔG. These sequences

should theoretically result in the best sensitivity, with the largest difference in signal between when the metabolite is present and absent.

As a test of this approach, a series of Tyr1-Spinach aptamers were synthesized and tested. Seven sequences were selected from those predicted to have the maximum ΔΔG; that is, each would exhibit the biggest difference in fluorescence compared when tyrosine is present or absent. However, the absolute ΔG of the dye binding in these sequences varied from slightly negative to the most negative predicted (Fig 4-6).

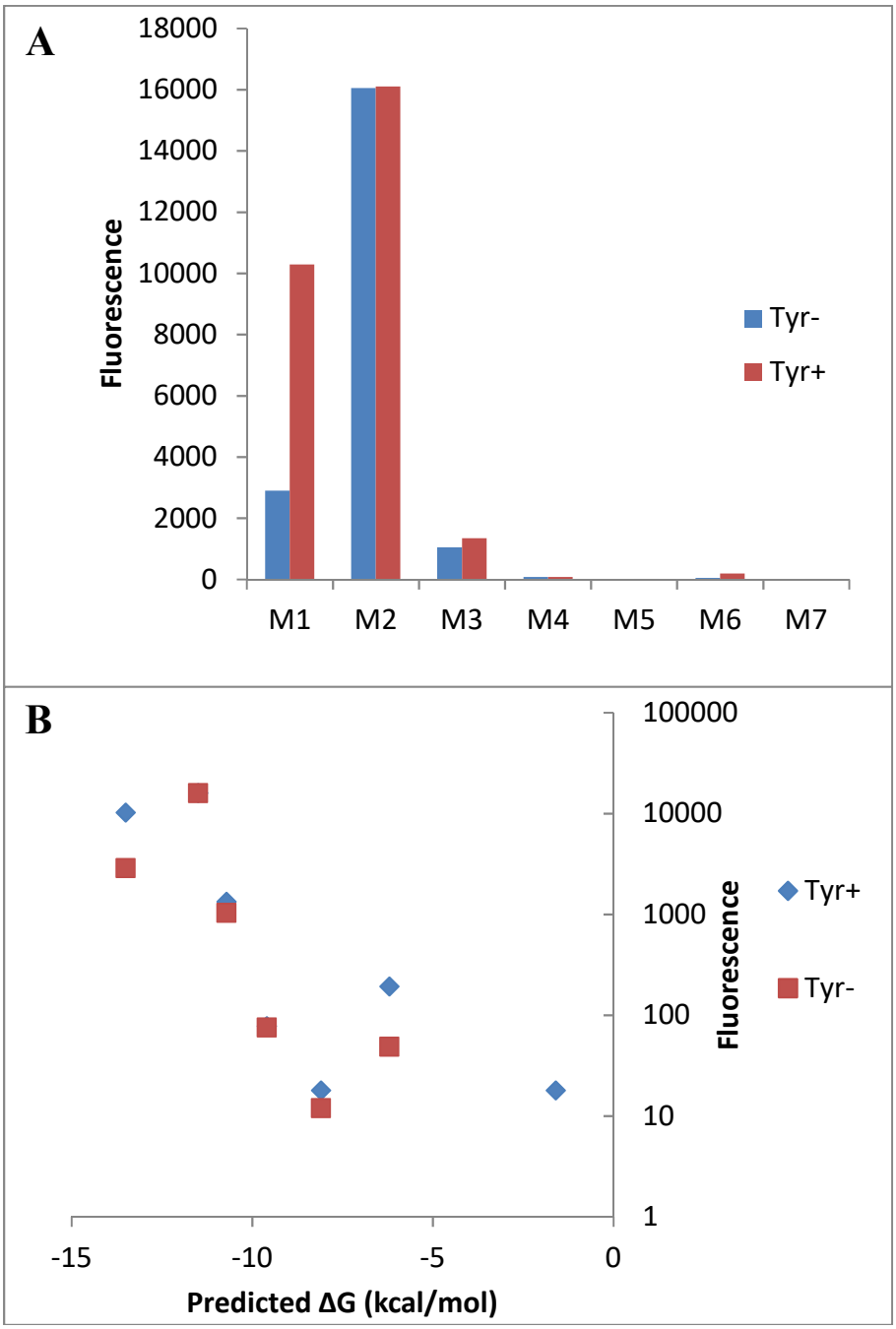
Figure 4-6: Predicted ΔG of binding of seven Spinach2-Tyr1 sequences optimized for sensitivity to tyrosine



Those with the more negative predicted ΔG should theoretically result in a higher signal, while those with the more positive values should result in a lower signal. All were predicted to have the same, maximal, ΔΔG, such that there should be a large difference in signal when tyrosine is present versus when it is absent.

Each of these sequences were next synthesized and tested with no tyrosine present and 1 mM tyrosine added (Fig 4-7). As predicted, those with the more negative ΔG of dye binding exhibited higher fluorescence. Furthermore, tyrosine increased this fluorescence in every case, although the difference was more pronounced in some compared to others.

Figure 4-7: Fluorescence values with and without tyrosine from optimized series of Spinach2-Tyr1 sequences



A) Fluorescence measurement of each aptamer with and without tyrosine present. **B)** these same measurements plotted against the predicted ΔG value from Fig 4-6, and demonstrates that, as predicted, the relationship is nearly linear on a log scale.

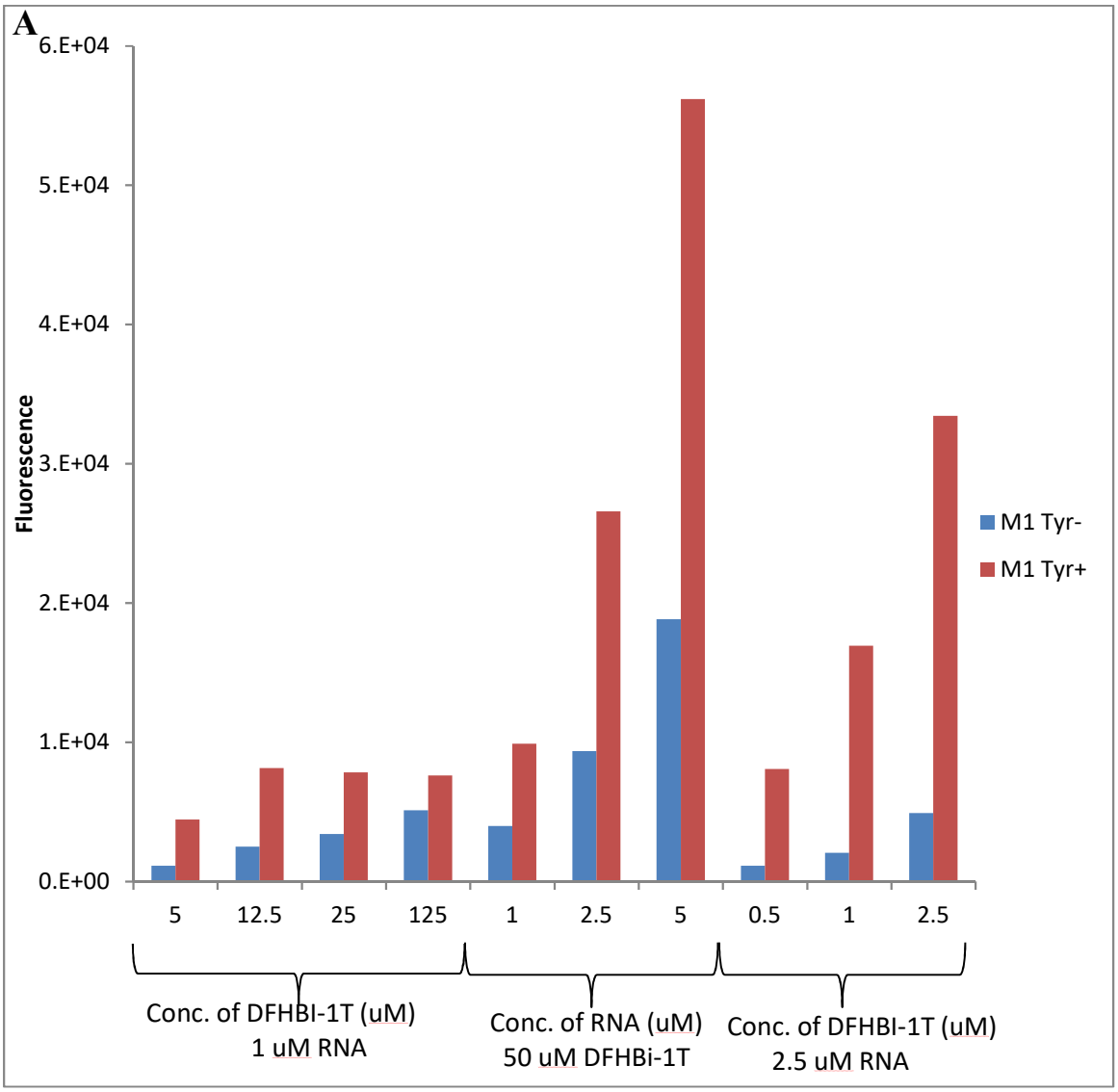
From this series, one sequence, called Tyr1M1, was selected for further experiments. It was predicted to have a very negative ΔG, and accordingly was one of the brightest aptamers tested. Further, it had the highest ratio of fluorescence between samples with tyrosine compared to those lacking tyrosine, with a ~3.5-fold increase.

4.3.3 Optimizing Biosensor Sensitivity

Initial experiments comparing various Tyrosine-Spinach aptamer sequences showed that the best observed resulted in a ~3.5-fold signal ratio between negative and positive controls (0 and 1 mM tyrosine respectively). Next, it was investigated if this sensitivity could be increased through optimizing assay conditions. Further experiments were performed to determine the effects of the concentration of dye and RNA, as well as the ratio between them, to optimize for detection of aromatic amino acids. A thermodynamic model of this process would predict that reducing the dye concentration substantially would increase this ratio.

Experimental results investigating this found the concentration of RNA had a significant effect on overall signal strength, while the ratio between dye and RNA impacted the signal ratio, both of which matched predictions (Fig 4-8a). Importantly, reducing the concentration of dye vastly improved the signal differentiation between negative and positive controls, with an optimal molar ratio of 2.5:1 resulting in an ~8-fold increase in fluorescence, again matching predictions (Fig 4-8b).

Figure 4-8: Fluorescence of Spinach2-Tyr1M1 at various conditions of dye and RNA concentrations



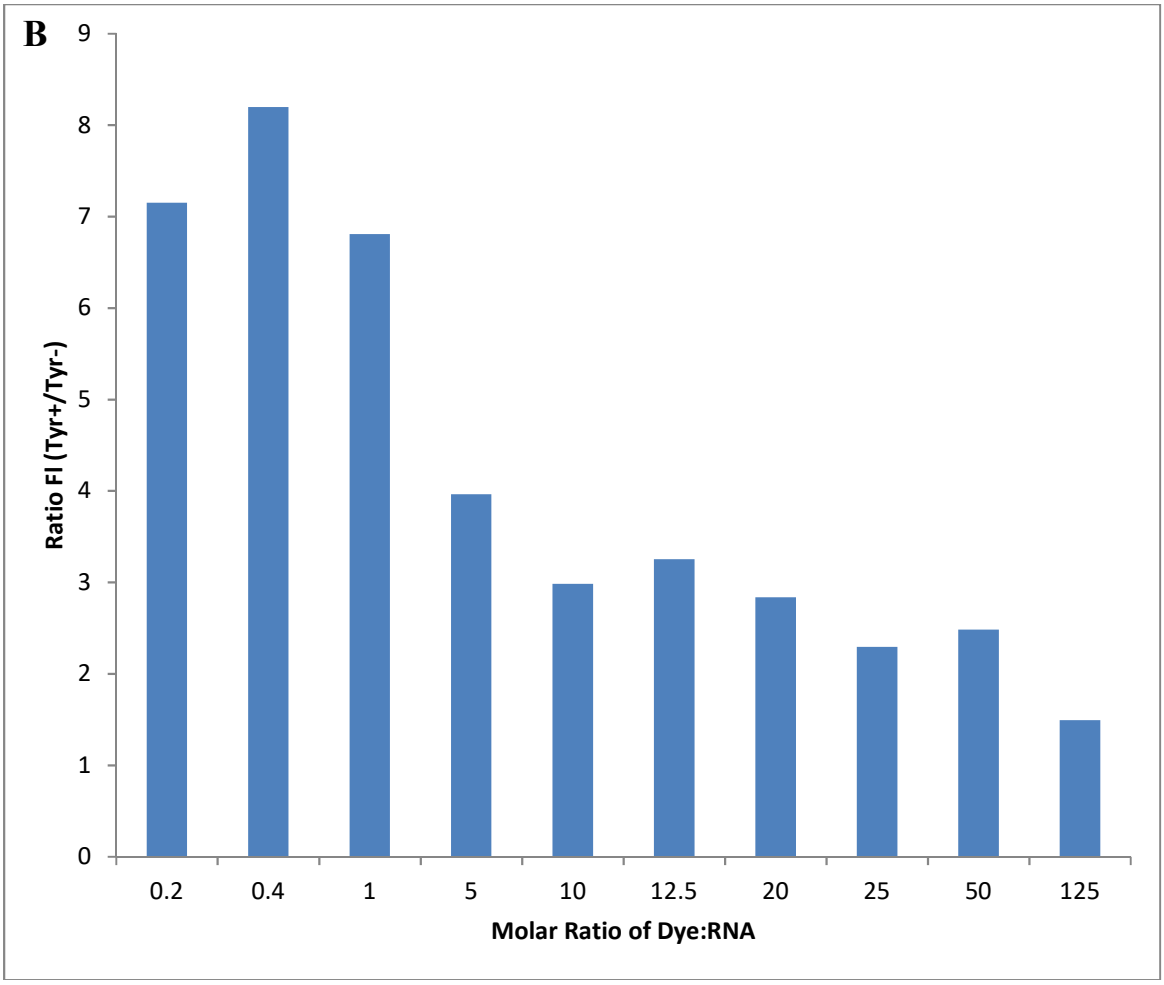


Figure 4-8 (continued): A) By varying both dye (DFHBI-1T) and RNA concentration one at a time, it could be determined that increasing RNA leads to a higher signal, while reducing the dye:RNA ratio leads to improved sensitivity. **B)** the same data, but with the ratio between the signal with tyrosine to the signal without tyrosine, as a function of dye:RNA ratio. Again, reducing this ratio leads to significantly higher sensitivity.

These results are very encouraging for the goal of utilizing these biosensors in a high-throughput screen. While the goal of any biosensor is to exhibit the highest sensitivity possible, it is often difficult to "tune" this sensitivity without redesigning the entire system. With the Spinach-based aptamer systems, the sensitivity can be tuned to the necessary level easily, by simply varying the ratio of dye and RNA added.

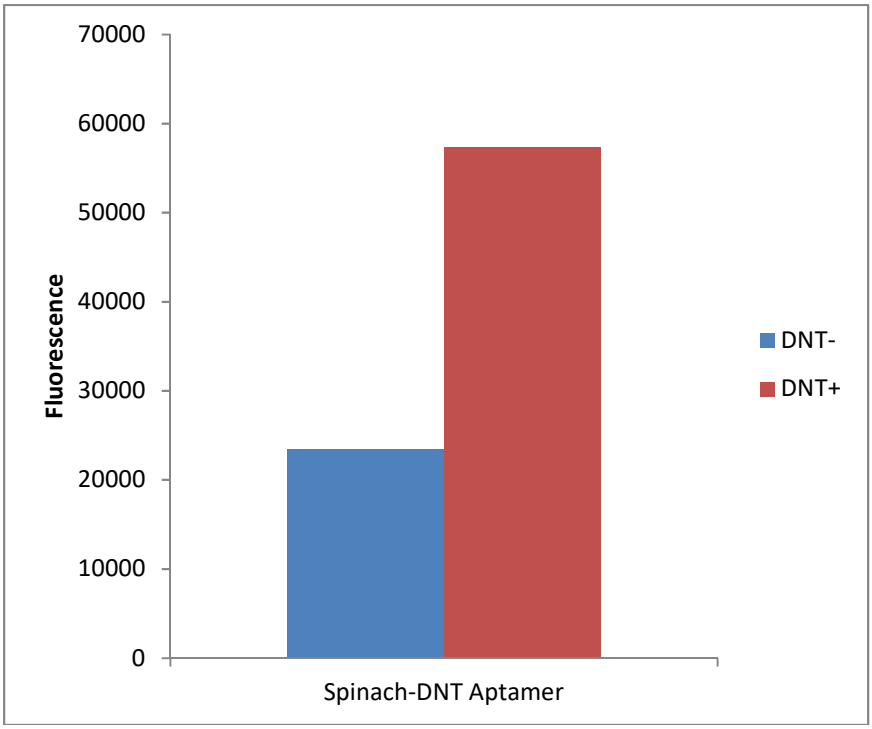
Furthermore, the desired signal can be amplified or reduced easily, by adding higher or lower concentrations of RNA.

4.3.4 Expanding Spinach-based Biosensor Modeling to Alternative Aptamers

Next, this model was expanded to alternative aptamers to demonstrate the generalizability of creating novel Spinach-based biosensors. While several metabolite-binding aptamers have been previously incorporated into the Spinach system, the linker regions of each were always tested by trial and error [133, 139], which is inconvenient for the rapid generation of a new biosensor. To validate the utility of the modeling approach used successfully to improve Tyr1-Spinach, it was next applied to a new aptamer that had never yet been incorporated to Spinach.

A novel aptamer which binds 2,4-dinitrotoluene (DNT), which was developed in the Salis lab but has not yet been published, was tested using the same linker modeling approach. A single sequence was selected for synthesis, based on the maximum predicted differential in ΔG between the metabolite-bound state compared to the unbound state. This sequence was synthesized and tested with and without 1mM DNT present. As predicted, the Spinach2-DNT fluoresced in the presence of DNT, with a 2.4-fold increase over background (Fig 4-9). This success demonstrates the power of the Spinach-linker modeling approach; without any expensive and inconvenient trial and error, a Spinach-DNT sensor was created that exhibits metabolite-dependent fluorescence. This again suggests such a system could be expanded to any functional aptamer to rapidly create a functional biosensor.

Figure 4-9: Fluorescence of Spinach-DNT aptamer with optimized linker region



This aptamer, predicted to have the optimal sequence for both brightness and sensitivity, was synthesized and tested with and without 1 mM DNT. It exhibited a 2.4-fold increase in fluorescence.

4.4 Discussion and Conclusions

While fluorescent biosensors have been used extensively in biotechnological applications, it remains difficult and time-consuming to create novel variations for a new target, which is critical in directed evolution of production phenotypes. To address this, we take advantage of RNA aptamer sensors linked to "Spinach", an RNA molecule capable of inducible fluorescence dependent on target metabolite concentration. Such sensors can act as transducers, linking the *in vitro* concentration of the metabolite of interest to fluorescence.

Here we greatly simplify and improve the process of creating novel Spinach-based biosensors. Instead of relying on a process of trial-and-error, we model predicted free energies to generate sequences that optimally link metabolite-binding aptamers to the base Spinach structure. This is demonstrated by optimizing Spinach-Tyr1, an aromatic amino acid sensitive biosensor, to vastly increase its sensitivity and signal brightness. It was also used to generate a completely novel biosensor using a brand new DNT-binding aptamer; a sensitive biosensor was created on the first try, without expensive and time-consuming synthesis of multiple variations.

With the success of the Spinach aptamer model thus far, this modeling approach can be expanded to even more complex Spinach-based biosensors. One possible next step will be to expand the model functionality by using it to calculate optimal sequences of multi-domain aptamers, which can bind two or more small molecules. By once again optimizing the linking sequence between each domain to result in conditional folding, it should be possible to make fluorescence sensitive to the presence of each molecule, and tune this sensitivity as desired. If successful, this would result in a multi-input, tunable biosensor system that can be expanded to a wide variety of target molecules and proteins.

Chapter 5 - Utilizing RNA Aptamers In Droplets (RAPID) Screening to Identify Production Phenotypes

5.1 Chapter Summary

A wide range of microorganisms have been effectively turned into living foundries through metabolic engineering and synthetic biology for the production of high value chemicals from renewable resources. However, identifying the phenotype of individual cell variants remains the main bottleneck to making rapid strain improvements. Here we demonstrate a novel high throughput microfluidic droplet screening approach using RNA aptamer biosensors, called RNA Aptamer In Droplet (RAPID) Screening. We show that a variety of small molecule- and protein-binding aptamers can detect the target of interest and act as a signal transducer, linking the concentration to a fluorescent signal, and furthermore that this system can be utilized in microfluidic droplets as a high-throughput biosensor. We demonstrate the utility of these sensors to sort individual yeast cells for high-production/secretion phenotypes in microfluidic circuits. These proof-of-concept applications can potentially be expanded to screen large microbial libraries for a wide variety of desirable targets.

5.2 Introduction

Synthetic biology embraces a design-build-test cycle to establish novel phenotypes. The benefit of this expanded design and build capacity can only be realized through improvements to the downstream test modality. In this regard, rapid screening of phenotypes (especially for the case of metabolite production) are required to reach parity with the $>10^9$ variant per week rate at which strains can be built through techniques including directed evolution. However, traditional metabolite detection approaches such as GC, HPLC, and MS are far too restrictive in throughput to accomplish this goal

without brute-force parallelization and high cost. As alternatives, biosensors may be used to link the concentration of a target molecule to fluorescence or a selectable trait such as antibiotic resistance, thus enabling a more rapid and quantitative detection [131, 132]. However, the development of such biosensors is difficult and requires host cell modifications that are likewise under positive selection pressure. In most cases, the product of interest will be secreted. This reality provides an additional challenge for any screening process that requires an intrinsic linkage between genotype and phenotype.

Droplet microfluidics relies on the interaction of immiscible liquids and can be used to create isolated picoliter sized reaction vessels in which cells can be encapsulated [140-143]. Recently, it has been demonstrated that yeast cells can be encapsulated and assayed individually in such droplets [79, 144]. Sorting capacity for these systems can reach as high as 10^8 droplets per hour [145]. In this regard, microdroplet enabled screening can address limitations in the test modality by physically entrapping genotype and phenotype - thus merging the benefits of single-cell analysis and high-throughput processing [144, 146]. Despite these advances, a remaining challenge precluding high-throughput screening is the capacity to transduce secreted product concentration into a detectable signal. As an alternative to protein-based biosensors, which may have difficulty accurately detecting secreted products, RNA aptamers have been explored for their potential to enable high-throughput screening. Specifically, RNA aptamers linked to "Spinach", a previously reported RNA molecule capable of inducible fluorescence, can be used to metabolite concentration into fluorescence [133]. Hundreds of RNA aptamers have reported in the literature with diverse binding kinetics to a host of analytes including small molecules and proteins [134]. This is especially useful when the Spinach-based biosensors are optimized for sensitivity using modeling techniques, as discussed in Chapter 4.

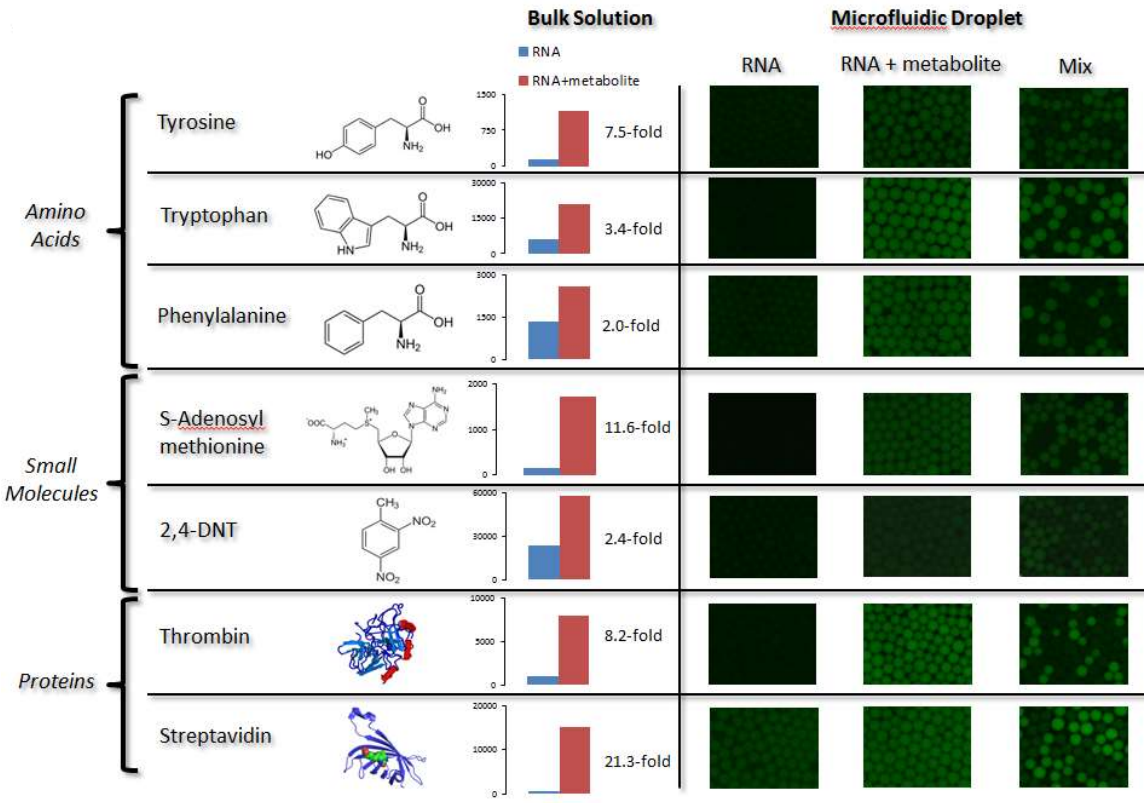
To demonstrate the capacity for high-throughput detection of secreted products in yeast, we develop and approach for RNA APtamer in Droplets, or "RAPID". In this scheme, we demonstrate that metabolites produced by yeast cultivated in microdroplets can be detected using Spinach-linked aptamters and sorted using a dielectrophretic mirofluidic device. Such an approach can be used to rapidly sort for improved variants in large libraries. Here we demonstrate a proof-of-concept application for engineering the yeast *Saccharomyces cerevisiae* for the overproduction of tyrosine. We demonstrate the isolation of significantly improved variants by using this high-throughput, generalizable screening modality that is not reliant on growth-based selection, chemical assays, or exogenous enzymes. Thus, this work establishes a generalizable high-throughput screening method applicable for microbe engineering and synthetic biology.

5.3 RESULTS

5.3.1 Importing RNA Aptamers to Microfluidic Droplets

Previously, RNA aptamers linked to "Spinach", a previously constructed RNA molecule capable of inducible fluorescence, have been used as sensors of metabolite concentration [133]. By inserting other aptamers into one of the stem loops of the Spinach aptamer, proper folding (and thus fluorescence) is made dependent on the presence and concentration of the second aptamers binding target. This has been demonstrated for a wide range of targets, including amino acids, nucleotides, and even proteins [134]. In order to co-opt this system as a novel high-throughput screening method, the capability of Spinach-based biosensors was extended to microfluidic droplets (Fig 5-1).

Figure 5-1: Spinach-based aptamer biosensors in droplets



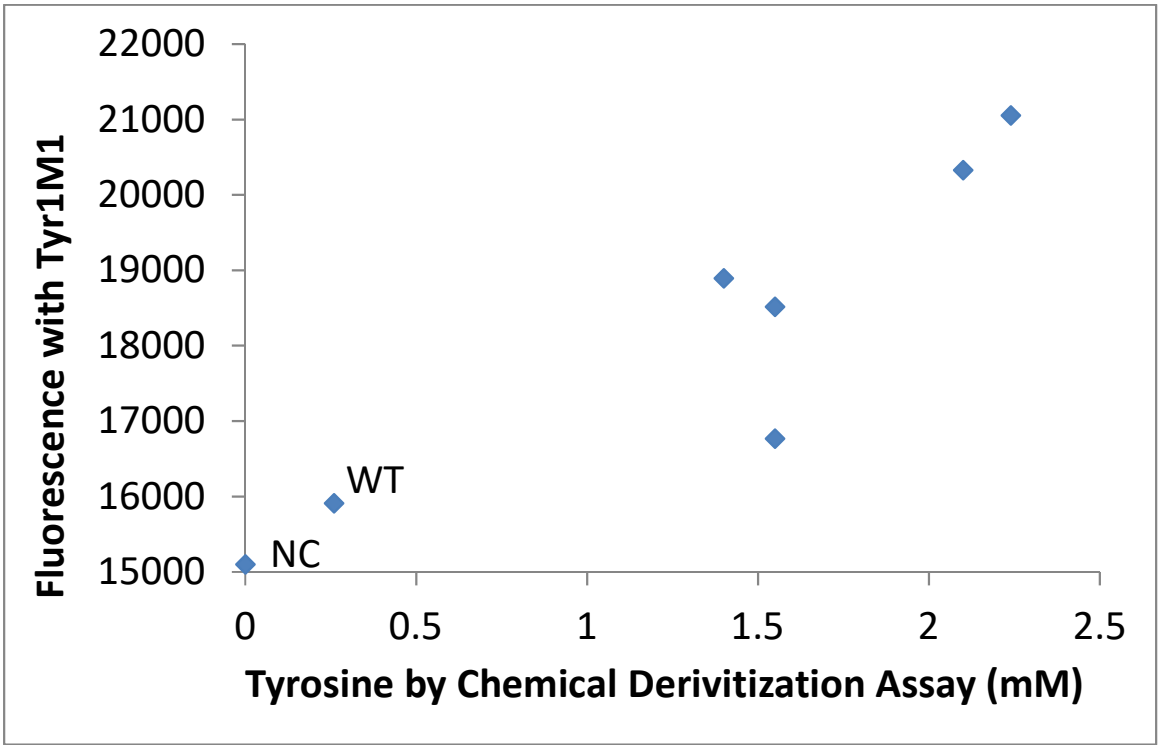
A diverse range of Spinach-based aptamers, binding to amino acids, other small molecules, or proteins, function as biosensors both in bulk solution and in microfluidic droplets.

Spinach-based biosensors for a wide range of targets, including novel targets like tyrosine and DNT, were created and tested, and then imported into microfluidic droplets. In each case, it was shown that the inducible fluorescence was maintained in a microfluidic droplet environment. These results confirm that a variety of aptamer biosensors can be utilized in conjunction with microfluidic droplets to create a novel high-throughput screening platform.

5.3.2 Validating Spinach-based Biosensors in Conjunction with Yeast

First, the most robust tyrosine-binding aptamer designed, Tyr1M1, was tested using the supernatant of yeast cultures after centrifugation, to approximate the conditions inside a microfluidic droplet with a single cell. Multiple strains engineered to varied levels of aromatic amino acid production were tested. These were independently first precultured in complete media, then washed and resuspended in a minimal media lacking aromatic amino acids. Several strains were tested, including one previously engineered to overproduce aromatic amino acids [4], along with three novel evolved strains that have been shown to overproduce even higher levels of aromatic amino acids (unpublished work). After incubation with the Tyr1M1 aptamer, fluorescence was measured (Fig 5-2). Each of the samples sourced from overproducing strains all did in fact result in a higher signal compared to that of the wild-type strain. Furthermore, when the WT sample was spiked to 1 mM tyrosine, the signal matched the positive control of 1 mM tyrosine, confirming that the aptamer functions correctly when mixed with media in which yeast cells have been incubated.

Figure 5-2: Tyrosine concentration in cell culture supernatants, tested either by Tyr1M1 aptamer or by chemical derivitization assay



Each point represents a different strain of yeast, with the wild-type labeled "WT" and a negative control solution labeled "NC". A strong positive correlation demonstrates that the Tyr1M1 can accurately detect higher tyrosine that has been produced and secreted into the media of growing cells.

5.3.3 Encapsulation and Screening of Tyrosine-Overproducing Yeast Strains

Next, RAPID was used to evolve a tyrosine overproduction phenotype by enabling the detection of high-producing and –secreting variants. Production of tyrosine is subject to tight regulatory control in yeast, and thus compounds derived from the aromatic amino acid pathway can suffer from poor yield. Molecules derived from the

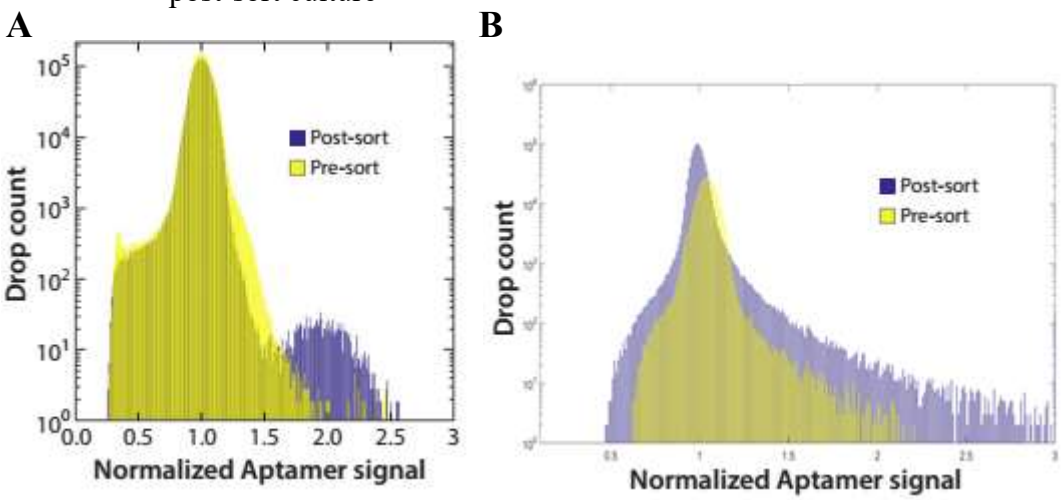
shikimate pathway (including tyrosine) have multiple uses as polymer precursors, nutritional supplements, and therapeutic agents [135].

While the RAPID Screening method could theoretically be applied to any library of mutants, we first explored mutagenizing a single biosynthetic enzyme: the DAHP synthase gene *aro4*. This enzyme has been well characterized, and it is known to be regulated through feedback inhibition by tyrosine. Furthermore, it was shown that such feedback inhibition can be removed through several point mutations previously reported, including the single point-mutation of K229L [147]; strains expressing this mutant have been shown to produce and secrete elevated levels of tyrosine [148]. Two libraries were made for the evolution of this gene: one was derived from the wild-type gene and one from the K229L mutant. To increase the chances of finding a beneficial mutant, the library derived from wild-type *aro4* used targeted mutagenesis to induce mutations only in the region of the protein known to control feedback inhibition, resulting in a library called ARO4-Reg. By investigating both libraries, it was hoped to replicate the discovery of novel feedback-reduced aro4 variants as well as to improve the current best variant known. Importantly, the same screening method could be used to screen libraries made by mutagenizing other genes, pathways, transcription factors, or even the entire genome.

These libraries were then screened in droplets using the Spinach-Tyr1M1 aptamer; the purified RNA was added to the culture media just prior to encapsulation, such that it could detect the concentration of secreted tyrosine from each cell (see Methods). After an incubation period, the droplets were sorted to isolate the top fluorescent, by gating to isolate the brightest ~0.1% of droplets. The sorted populations were then recovered in fresh media and grown to stationary phase, and then re-encapsulated as before to quantify the sorting efficiency. The relative tyrosine production

of this population was again measured by aptamer-mediated fluorescence compared to the original library, clearly demonstrating the relative enrichment (Fig 5-3).

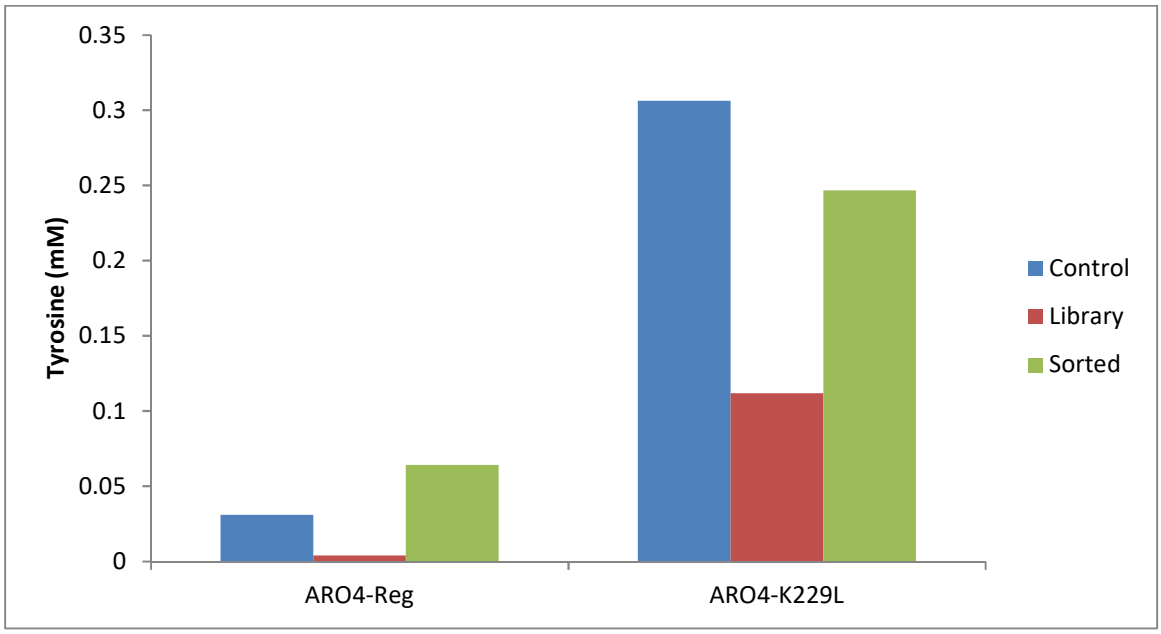
Figure 5-3: Distribution of aptamer-mediated fluorescence in pre-sort library compared to post-sort culture



A) is the library with *aro4* mutagenized in the regulatory region (ARO4-Reg), **B)** is a library of mutagenized $aro4_{K229L}$. Both demonstrate significant enrichment.

In addition, the bulk sorted cultures were tested for tyrosine production using a derivatization assay [149] (see Methods). Cultures from both sorts exhibited higher levels of tyrosine production compared to the original library (Fig 5-4). Thus, the sorting efficiency was demonstrated by two orthogonal measurements, confirming the aptamer-mediated fluorescence functioned as intended.

Figure 5-4: Secreted tyrosine concentrations in bulk sorted cultures for two libraries, ARO4-Reg and ARO4-K229L



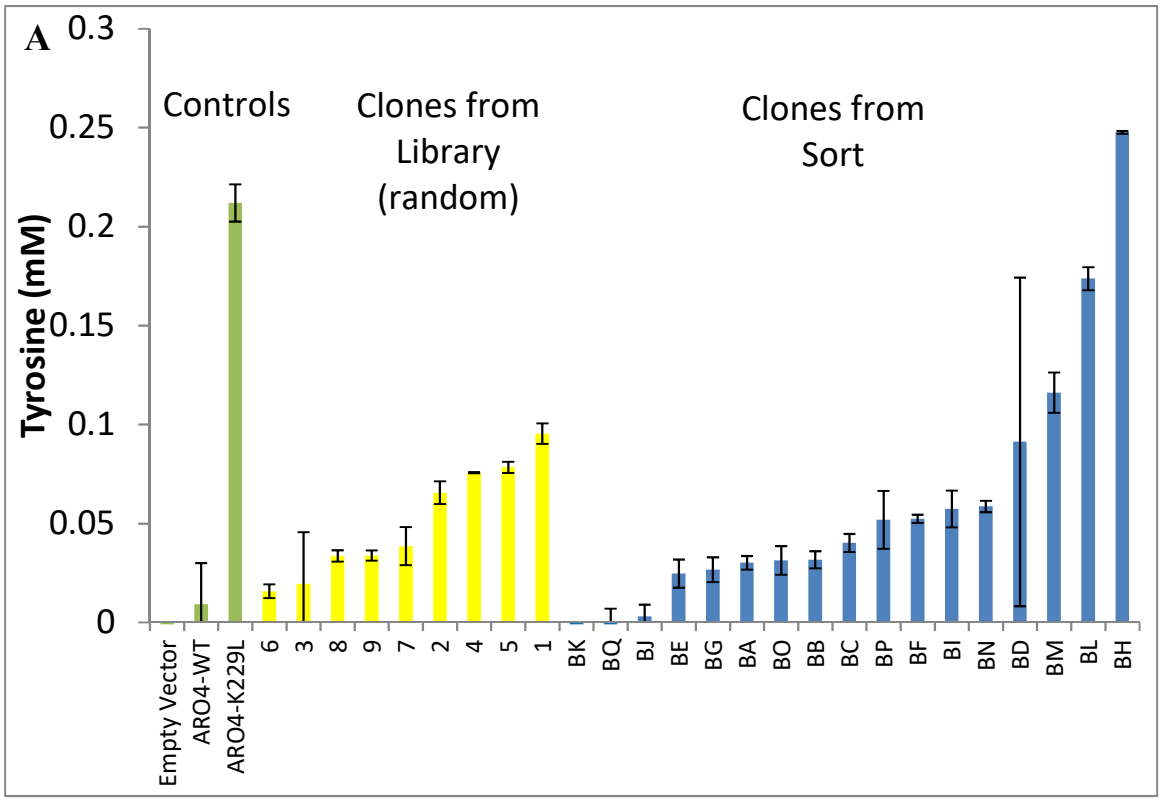
In each, the "Control" bar indicates a culture expressing either *aro4* or $aro4_{K229L}$. The "Library" bar is a bulk culture of the pre-sorted library; these are significantly lower, presumably due to the presence of empty vector background and deleterious mutations. The "Sorted" bar is a bulk culture of the post-sorted library, recovered in fresh media. In both libraries, the sorted bulk is significantly enriched in tyrosine producers.

5.3.4 Validation of Tyrosine Over-Production from Sorted Mutants

After the initial sort, individual mutants were isolated and re-transformed to a fresh strain of yeast, in order to eliminate any strain adaptations that arose during the microfluidic encapsulation and sorting. These strains were then grown and incubated, and the media was assayed for secreted tyrosine concentration using a derivatization assay [149] (see Chapter 8: Materials and Methods).

For the *ARO4*$_{K229L}$ library, the collection of individual clones isolated from the sort were significantly enriched in tyrosine over-producers compared with clones containing randomly mutated sequences obtained from the unsorted library (Fig 5-5b, $p<0.05$, Mann-Whitney U test).

Figure 5-5: Tyrosine secretion from clones expressing *aro4* mutants



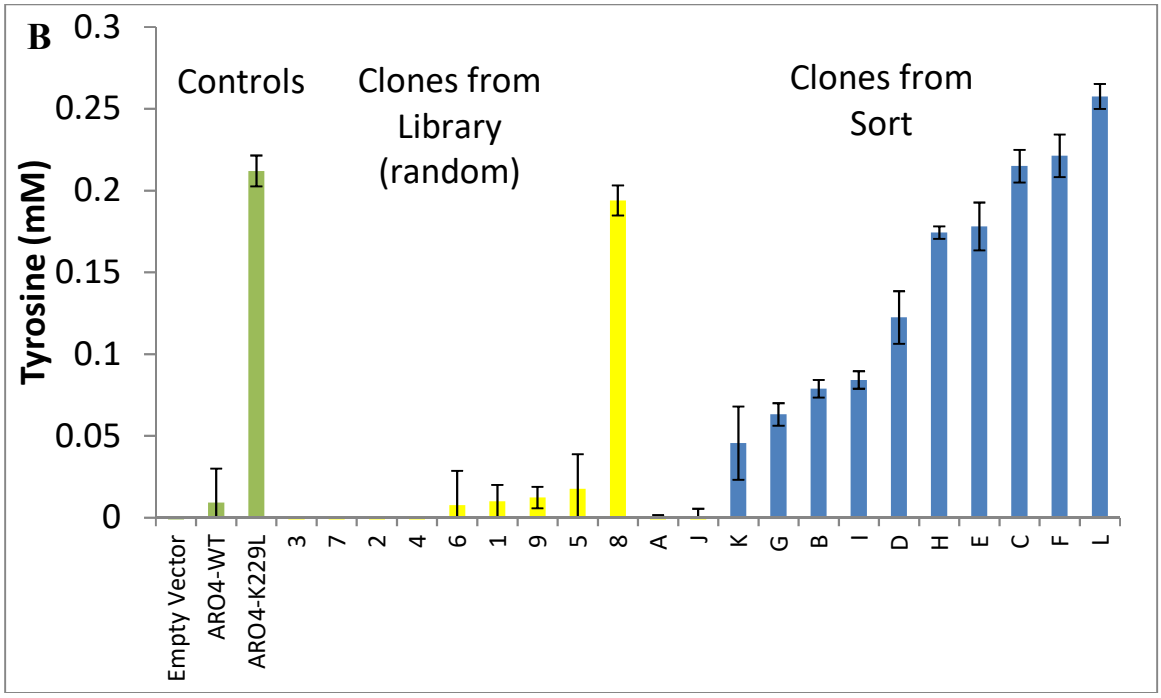
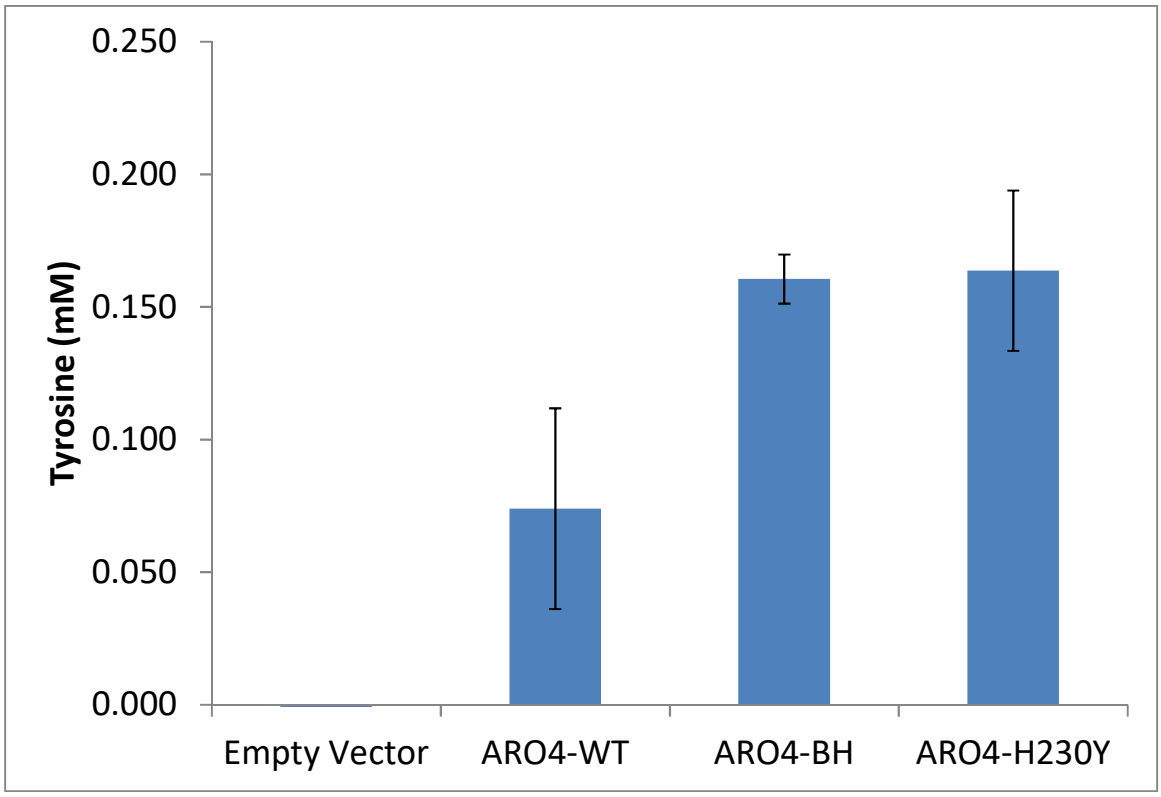


Figure 5-5 (continued): Control vectors (green), randomly mutated (yellow), or isolated from sorted culture (blue). **A)** contains clones from ARO4-Reg library, and **B)** from ARO4-K229L Library

In addition, both libraries revealed novel variants that were significantly improved over their base sequence. In fact, the best from each library (ARO4-L and -BH, derived from *aro4*$_{K229L}$ and wild-type *aro4* respectively) both conferred improved tyrosine production compared to *aro4*$_{K229L}$, which until now has been the mutant which enabled the highest production of aromatic amino acids. Interestingly, both of these two novel variants contain mutations in the same region, which is known to control tyrosine feedback inhibition; this gives some clue as to the mechanism of mutation. ARO4-L has two coding point mutations in addition to the K229L scaffold from which it was derived: T46A and T207I. ARO4-BH has three coding mutations: H230Y, K252N, and V262I. H230Y is particularly interesting, since this is adjacent to the lysine at 229 known to have

significant effects. This mutation was tested independently, and was found to recapitulate the increased tyrosine expression (Fig 5-6).

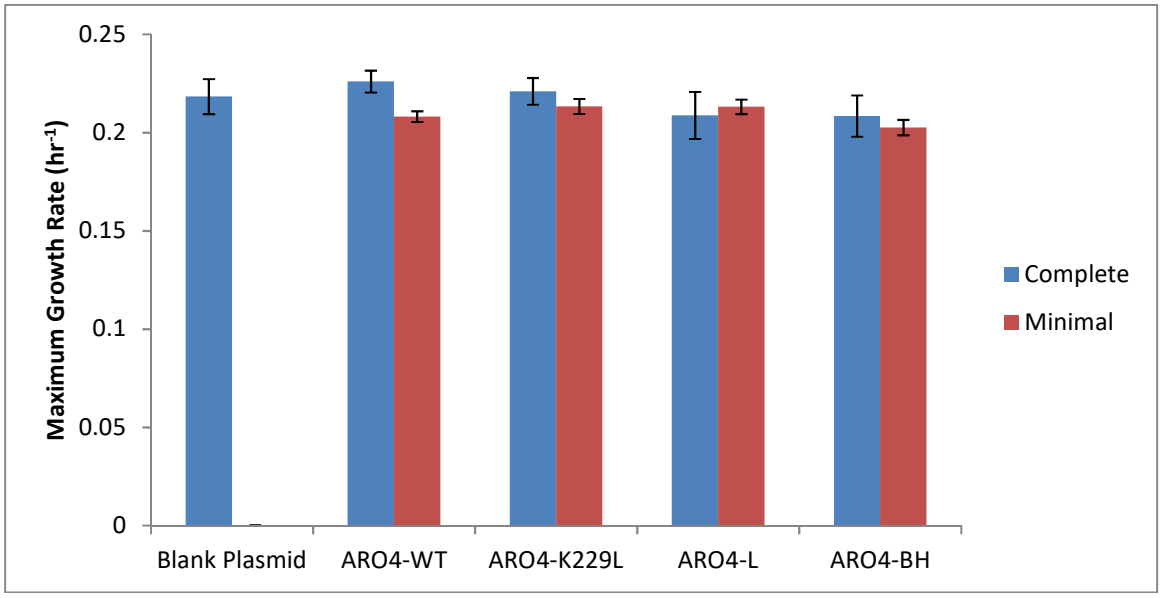
Figure 5-6: Tyrosine secretion from clones expressing *aro4* mutants derived from ARO4-BH



The mutation H230Y alone recapitulates the increase in production of the triple-mutant ARO4-BH.

Next, the growth rate of strains expressing these mutant *aro4* genes was characterized, in both complete media and minimal media lacking aromatic amino acids. Importantly, the growth rates of these strains were not significantly different from cells expressing the wild-type gene (Fig 5-7). This demonstrates that such a phenotype could not have been identified through growth based selection; high-throughput screening was necessary.

Figure 5-7: Growth rates of strains expressing *aro4* variants



Blue bars represent complete YSC media, while red bars represent minimal media containing only glucose, auxotrophic amino acids, and YNB. There is no significant growth difference between cells expressing wild-type *aro4* and the high-producing mutant versions, either with or without aromatic amino acids present.

While further investigation of these novel mutations is certainly warranted, here they serve as an important proof-of-concept for this system. Never before has it been possible to screen *aro4* mutants in a high-throughput manner; all previous beneficial mutations relied on a known crystal structure and an understanding of protein dynamics [150], both of which are lacking for the vast majority of enzymes of interest. The success in identifying previously unknown beneficial *aro4* mutations using this screen definitively demonstrates that Spinach-based aptamers in droplets is a powerful tool for metabolic engineering and directed evolution.

5.4 Discussion and Conclusions

Here we have demonstrated that metabolites produced and secreted by yeast can be detected and sorted using Spinach-linked aptamers in conjunction with a dielectrophretic microfluidic device. In this, we have created the potential for a rapid, high-throughput screen that can easily enable sorting of large libraries. This proof-of-concept system further expands the potential for directed evolution techniques to evolve an even wider array of phenotypes in a rapid and high-throughput process.

Importantly, by demonstrating and optimizing an aptamer-based selection screen to rapidly detect metabolite concentrations *in vitro*, we have further shown that this platform is potentially generalizable to a vast array of targets. Challenges do remain, especially if an aptamer has not yet been developed for a target analyte or if the aptamer has not yet been linked to Spinach. These processes are almost certainly target specific, and may be more difficult for some than for others. However, we demonstrate that it is quite possible to quickly import and adopt a novel aptamer into this system. Further, by importing multiple aptamer sensors to a microfluidic system for the first time, we were able to create a novel and powerful platform for screening finally commensurate with possible diversity generation.

Chapter 6 - Creation of Novel Pathway Engineering Tools Using Viral 2A Cleavage Sites

6.1 Chapter Summary

In directed evolution of pathways, it is difficult or cumbersome to express and evolve multiple enzymes simultaneously, requiring that a significant proportion of synthesized and mutagenized DNA consist of regulatory elements. Viral 2A sequences enable multi-protein expression and mutagenization in a single cassette by encoding co-translation polypeptide cleavage. By inserting these sequences between pathway enzymes, it is possible to evolve an entire pathway simultaneously, as was done for the xylose catabolic pathway in Chapter 2. Here we engineer these sites by creating a novel, secretion tag-based high-throughput assay allowing rapid quantification of 2A efficiency. Further, we use this assay to discover novel sequence elements that significantly improve cleavage efficiency. However, challenges remain in successfully applying 2A sites for expression of metabolic engineering relevant contexts. In many cases, fusion proteins retain activity, rendering the 2A cleavage superfluous; in others, the remnant 2A peptide post-cleavage can affect activity. The characterization of these novel high-activity 2A sites offers a powerful new tool for future pathway engineering efforts.

6.2 Introduction

In order to effectively engineer pathways of enzymes, novel tools need to be developed for the efficient simultaneous expression of multiple proteins. Commonly, heterologous pathways constructed in yeast use a promoter and terminator motif for each individual gene, often using multiple vectors. This can be disadvantageous for several reasons. First, while there are an expanding number of regulatory elements being characterized, there remains a small number that are desirable for repeat use, especially

those that result in strong or inducible expression. This can be especially true for non-model organisms in which only a handful of regulatory elements are well-characterized [151]. This often leads metabolic engineering efforts to repeatedly use the same promoters and terminators, resulting in a high probability of homologous recombination and thus genomic instability. Second, the use of multiple regulatory elements for each gene necessitates a large fraction of synthesized DNA to be non-coding. For extensive metabolic engineering applications, this can significantly increase the cost and difficulty of synthesis and transformation. Finally, in an evolutionary engineering approach, it is very difficult to co-evolve multiple genetic parts; often, each gene or part needs to be mutagenized individually, making it impossible to discover synergistic double mutations. While these limitations can be addressed partially with the use of fusion proteins, this is often unworkable, either due to protein folding complications or to non-localized function (eg cytoplasm vs mitochondria). Further, these limitations are not often known beforehand.

Here, we aim to engineer novel pathway engineering tools that allow co-expression of multiple proteins with a single promoter and terminator. Such a tool enables rapid and simple pathway engineering, while minimizing the sequence length needed. For evolutionary engineering of pathways, such tools would allow the mutagenesis of an entire multi-gene pathway, without concurrent mutagenesis of the regulatory regions, vastly increasing the likelihood of generating beneficial mutations in the coding region.

These novel tools are based on viral 2A sequences. These short sequences are about 20 amino acids in length and are found encoded in viral genomes to allow translation of single large polypeptides, which self-cleave into multiple smaller proteins. When a ribosome translates a 2A sequence, the nascent polypeptide quickly cleaves to

form a separate protein, while translation continues downstream to form a second protein [72]. In this way, 2A sequences allow a single promoter to drive expression of an open reading frame encoding multiple proteins. Furthermore, these sequences have been previously shown to be active in *Saccharomyces cerevisiae* [73, 74].

While such sequences have great potential as tools for metabolic and evolutionary engineering in yeast, several challenges remain. One significant issue is that polypeptide cleavage is not perfect; uncleaved polypeptide always remains present. If an aim requires co-expression of multiple proteins, each additional 2A site reduces the fraction of fully cleaved proteins. Further, this results in heterogeneity of expression, with various mixtures of cleaved and uncleaved polypeptides present in the cell.

Unfortunately, 2A activity is not well understood. Several unique 2A sequences have been characterized, but their cleavage activity is not uniform [74]. However, it is unknown which factors affect activity. While four 2A sequences active in yeast have a conserved active site including a C-terminal NPGP domain, the N-terminal sequence differs greatly between the four, in sequence and even in length. Previous tests in our lab have demonstrated that even codon usage affects activity; while P2A was shown to promote active cleavage in yeast, changing 22 of its 66 nucleotides with silent mutations resulted in a complete loss of activity [152].

A second challenge in using 2As for pathway engineering is the difficulties in using them to build multi-enzyme cassettes. Replicating 2A sequences multiple times leads to the same chance of homologous recombination instability discussed above, while at the same time is obviously disadvantageous to use less active 2A sequences simply to avoid homology. Since silent codon modification can also affect activity, there is no ideal option for the use of 2A sites in multi-gene pathways.

In an effort to engineer these tools for their use in pathway engineering, we worked to create novel or synthetic 2A sequences that are optimized for high activity in yeast. Furthermore, we aimed to identify multiple, sequentially diverse sequences that are uniform in high cleavage activity. Such a set of tools would significantly improve pathway engineering and the directed evolution of these pathways.

6.3 RESULTS

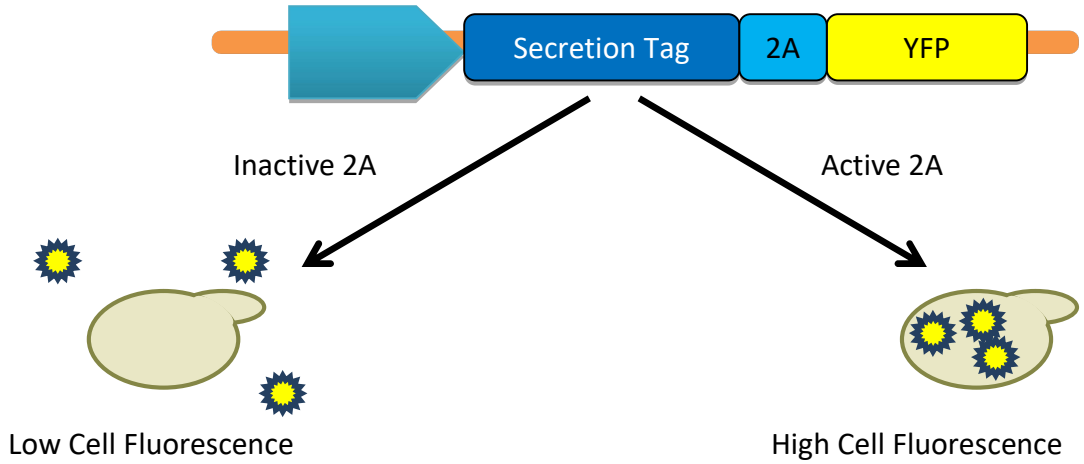
6.3.1 Creating Quantitative Assay of 2A Activity Using Secretion Tag

Previously, multiple 2A sites have been tested and characterized in *Saccharomyces cerevisiae*. For example, P2A, derived from porcine teschovirus-1, has been shown to be highly active compared to others [72]. However, there remains no easy way to quantify or directly compare different 2A sequences head-to-head; currently activity is evaluated by Western blot relative band intensities, which can be inaccurate.

To evaluate and compare 2A sites for pathway engineering applications, a novel assay was first developed using a secretion tag. In this assay, an N-terminal secretory signal was paired with a fluorescent protein [153]; with this signal active, the protein is secreted into the media, resulting in relatively low cell fluorescence as measured by flow cytometry. A 2A sequence can then be inserted between the secretory signal sequence and the fluorescent protein. If such a sequence is effective at cleavage, we expect the secretory signal will be removed and degraded, resulting in high cell fluorescence; otherwise, the entire protein will be secreted, resulting in low cellular fluorescence (Fig 6-1). Such an assay represents a significant improvement for quantifying 2A activity: it allows direct quantitative comparison of 2A sequence efficiency, and it allows rapid

turnaround to test novel sequences, relying only on the relatively high-throughput flow cytometry.

Figure 6-1: Schematic demonstrating how a secretion tagged yellow-fluorescent protein can be used to assay 2A activity

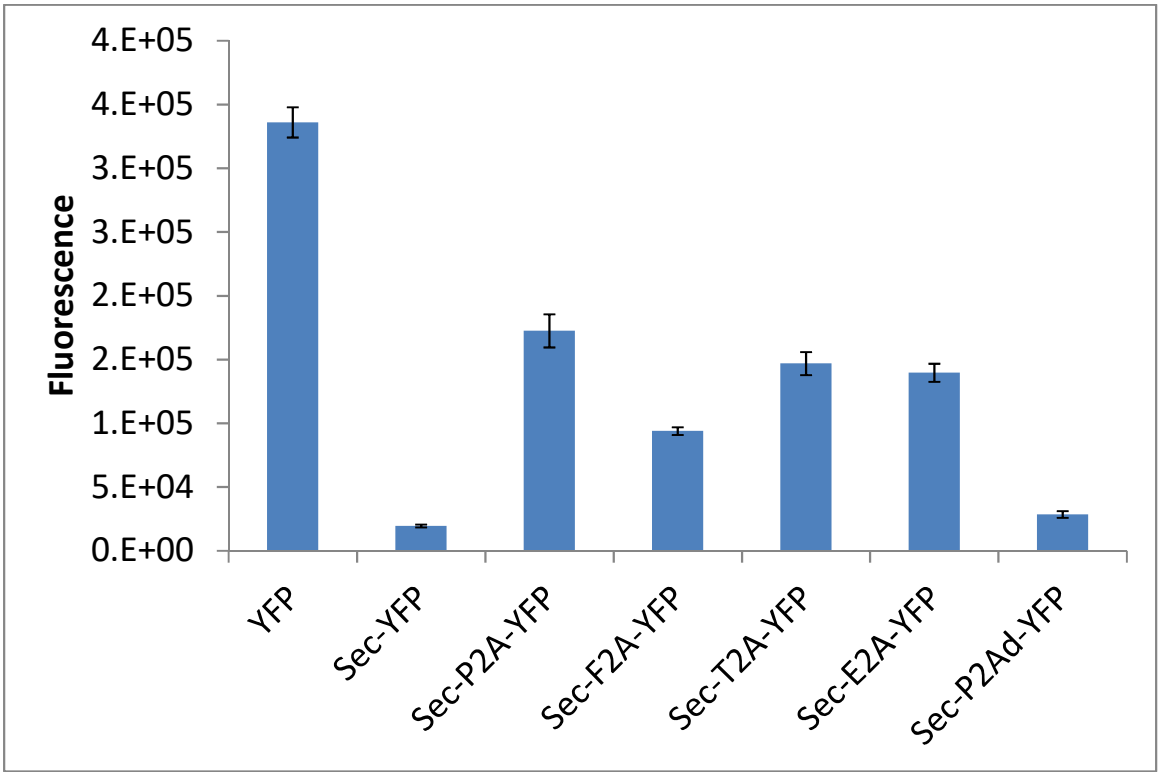


If the 2A is inactive, YFP will be secreted from the cell, resulting in low cell fluorescence (left). If the 2A is active, the tag will be cleaved during protein synthesis, and the YFP will be retained by the cell (right). Cell fluorescence can then be measured quickly and quantitatively using flow cytometry.

To demonstrate this assay's utility, four native 2A sequences were inserted (P2A, F2A, T2A, E2A), as well as a non-functional control made by mutating the active site of P2A (called P2Ad [154]). These plasmids were then transformed into yeast and assayed by flow cytometry (Fig 6-2). As expected, cells expressing the secretion tagged YFP exhibited low cell fluorescence, roughly 5% of the fluorescence of those cells expressing YFP with the same promoter. Inserting P2A between the secretion tag and the YFP gene, though, increased the fluorescence significantly, by ~9-fold. However, this value was only ~50% of the positive control, presumably due to only partial cleavage activity.

Furthermore, the other three native 2As tested had variable activity, which when rank ordered matched the same relative activities observed previously. Finally, the P2Ad construct matched the negative control, suggested that the 2A sequence itself did not materially affect fluorescence without active cleavage.

Figure 6-2: Cell fluorescence as measured by flow cytometry of cells expressing YFP with a TEF promoter



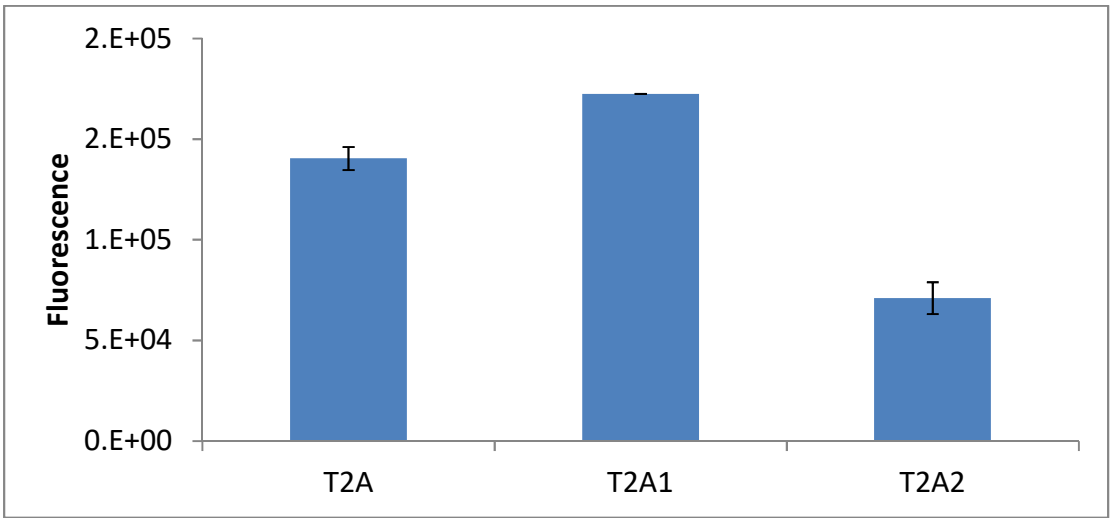
Sec-YFP is the same construct with the secretion tag added to the N-terminus of YFP, resulting in secretion of YFP and low cell fluorescence. The next five are the same constructs with various 2A sequences inserted between the secretion tag and YFP.

These results were very encouraging for several reasons. First, this assay was established as a quantitative and high-throughput way to assay 2A sequence activity. Second, the fact that the relative activities measured corresponded to the relative

activities previously observed by Western blot served to validate that the assay does in fact measure 2A cleavage. Finally, the fact that P2A exhibited only ~50% of the fluorescence of the positive control indicated that 2A cleavage is incomplete, even that of the most active sequence known. This implies that it may be possible to find a variant sequence that exhibits higher cleavage activity than any known native site.

As a quick demonstration of the utility of this assay, we next used it to investigate 2A sequences used in a previous metabolic engineering effort in yeast [74]. In this approach, investigators used two T2A sites to assemble a 3-enzyme pathway as one genetic cassette. To avoid homologous recombination, they changed the codon usage of the second T2A linker. While a Western blot confirmed cleavage was taken place, there was no easy way to quantitatively measure the effect codon usage had on the T2A activity. We synthesized these two sequences and inserted them into the secretion tag assay, and ran them along with the T2A sequence tested above, which had been codon optimized for *Saccharomyces* and was sequentially unique. Interestingly, we observed a significant decrease in cleavage activity by our assay with the codon-altered T2A sequence (Fig 3-3). This again confirms how critical a quantitative measure of 2A activity can be, as it remains unknown which factors, such as codon usage, can have significant effects.

Figure 6-3: Cell fluorescence as measured by flow cytometry of cells expressing the TEF-Sec-2A-YFP system



The codon optimized T2A (left), or two codon-altered version of the same sequence used in a previously published study (center and right).

6.3.2 Using Secretion Assay to Sort 2A Sequence Library for Novel High-Activity Elements

Using the secretion tag assay approach, putative 2A sites can be quickly tested for effectiveness using flow cytometry; higher cell fluorescence will indicate effective and efficient cleavage. This assay can also be used to find novel 2A sequences from a library of randomized sequences, since FACS makes it possible to sort out high fluorescent cells from a population. In a population of cells, each containing a randomized 2A sequence in the secretion tag construct, those exhibiting the highest fluorescence presumably harbor 2A sequences with the highest activities. Ideally, such a library would reveal both high activity sequences as well as novel, non-homologous sequences; both would be useful in pathway engineering efforts.

This library approach was undertaken with several methods. In the first, a library was assembled directly using yeast homologous recombination, by co-transforming an

insert amplified from a primer with degenerate nucleotides and a backbone with homology containing the secretion tag and fluorescent protein. Enough homology was included to ensure recombination, such that each cell would contain a unique 2A sequence but the same secretion tag and fluorescent protein gene. The 9 amino acids consisting of the active site of the 2A sequence was maintained in the homologous region, with 27 base pairs directly upstream randomized. When recombined, this would presumably result in a completely novel 2A sequence with the same conserved active site. This library contained ~1.2 x 10^5 unique variants, a tiny fraction of the 4^{27} possibilities (~$1.8x10^{16}$). While it could not be expected to identify the optimal sequence with such an approach, it was theorized that novel sequences dissimilar from the native sequences could be identified which exhibited activity; from these, further variations could be explored. The highest fluorescent cells were sorted from this population using FACS and plasmids were recovered and re-transformed into fresh strains for assaying.

Several sequences were isolated from this sort that exhibited significantly higher fluorescence that control 2A sequences, and comparable to a positive control with no secretion tag at all (Fig 6-4a). However, on analysis it was determined that these were not simply novel 2A sequences, but instead recombination events of two DNA insert pieces, resulting in 2A sequences about double the length of the native. Some of these had two complete 2A sequences concatenated, while others contained the complete T2A sequence with an incomplete fragment of 2A sequence combined upstream. While it could be expected intuitively that including two concatenated 2A sequences would result in high activity, it was not clear why those with an incomplete 2A sequence added to the upstream end of a complete 2A sequence, called “spacers” hereafter, would increase the activity. The “spacer” sequences of two such sequences, denoted d8 and d12, were cloned into a separate plasmid without a downstream 2A site, and it was shown that alone they

do not impart any cleavage activity (Fig 6-4b). When placed upstream of the complete T2A sequence, however, they do. These spacer elements were explored further, which is discussed in the next section (6.3.3).

Figure 6-4: 2A activity of library sort isolates, as measured by cell fluorescence from the TEF-Sec-2A-YFP assay system

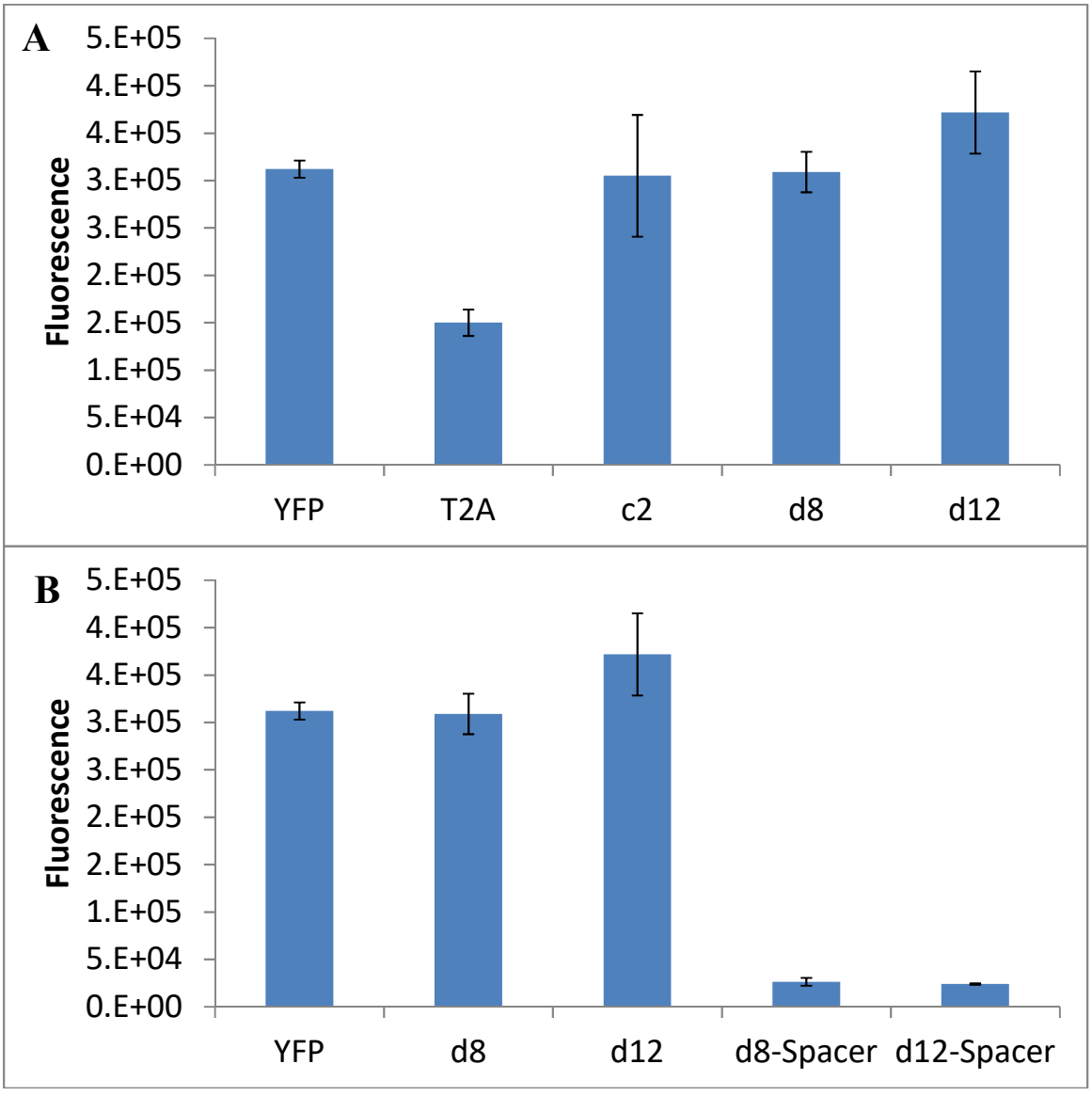
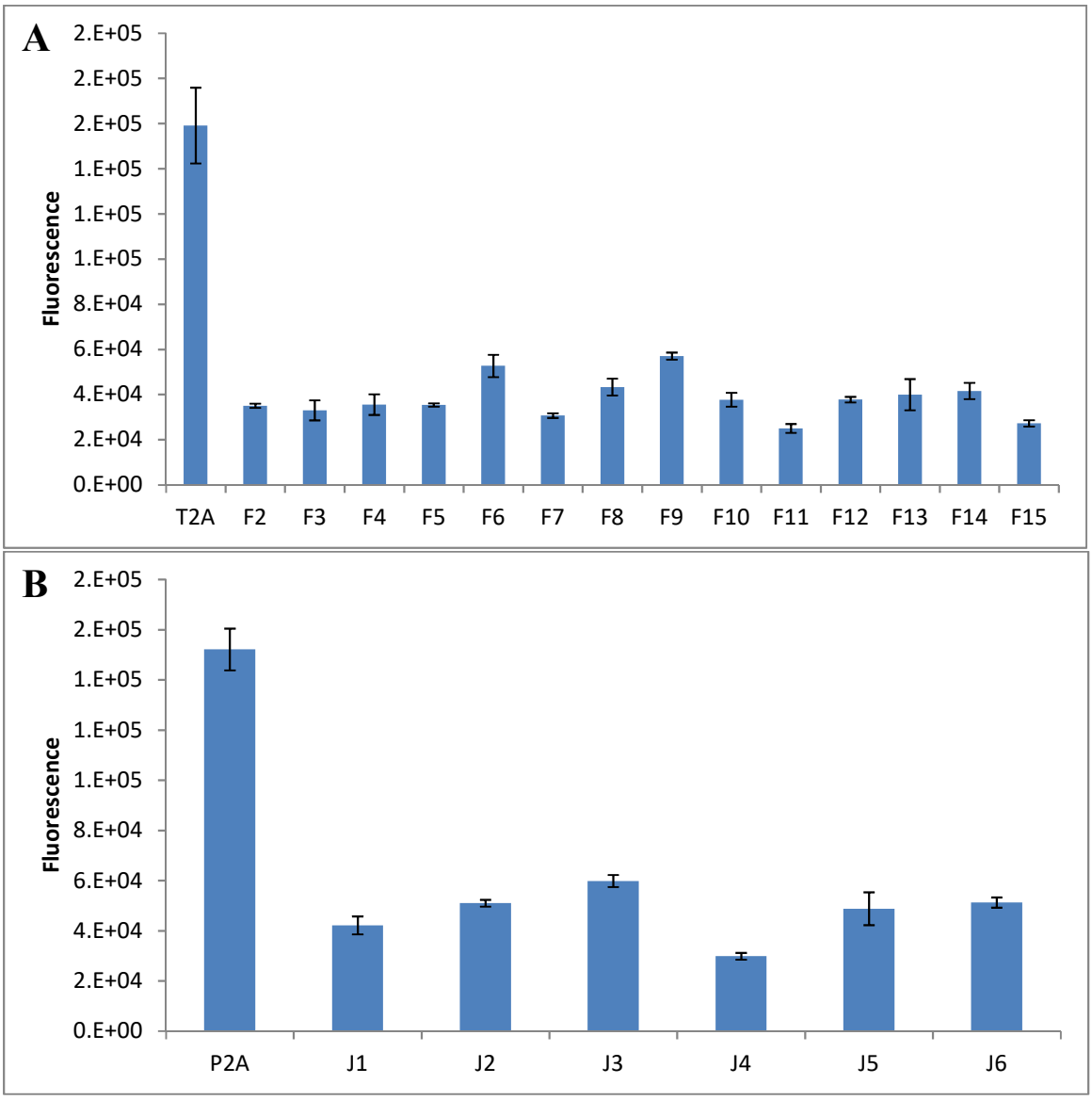


Figure 6-4 (continued): A) Leftmost is a control expressing TEF-YFP with no secretion tag as a positive control, with the next containing Sec-T2A-YFP as a control comparison to a native sequence. c2 is a novel sequence isolated from library, consisting of two concatenated full 2A sequences. d8 and d12 are two sequences isolated from library, consisting of an incomplete 2A "spacer" sequence upstream of T2A. All three exhibit extremely high cleavage activity, comparable to having no secretion tag at all. **B**) The first three points are controls, with d8-Spacer and d12-Spacer consisting of solely the incomplete 2A isolated from d8 and d12. These spacer elements exhibit no cleavage activity alone, suggesting the high activity of d8 and d12 is not due to multiple active sites.

Several additional libraries were then constructed with alternate cloning approaches in order to identify novel short 2A sequences as well. Three separate additional libraries were made and transformed using Gibson assembly, which would theoretically result in significantly reduced recombination of two or more inserts. These libraries (containing unique 8×10^5, 6×10^4 and 7×10^4 variants each) were each sorted using FACS to isolate the brightest 0.1% of cells. Next, plasmids were isolated from the sorted population and re-transformed. While several sequences exhibited some activity, none were comparable to the wild-type native 2As (Fig 6-5).

Figure 6-5: 2A activity of additional library sort isolates, as measured by cell fluorescence from the TEF-Sec-2A-YFP assay system



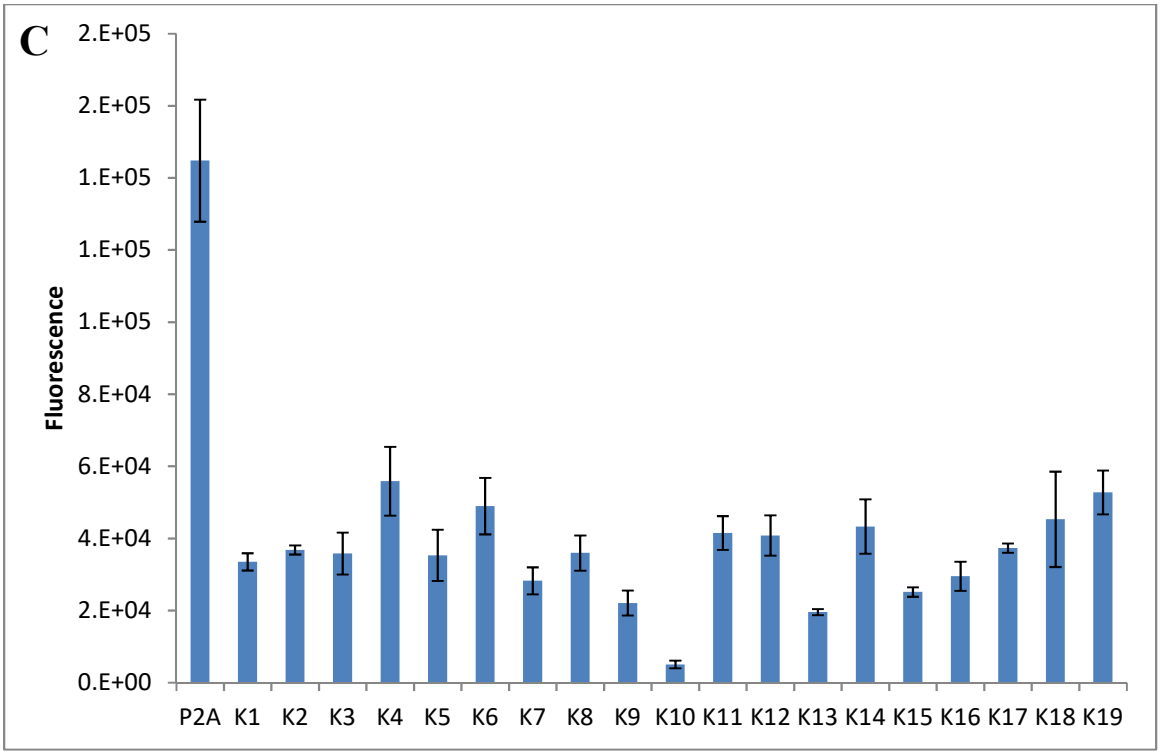


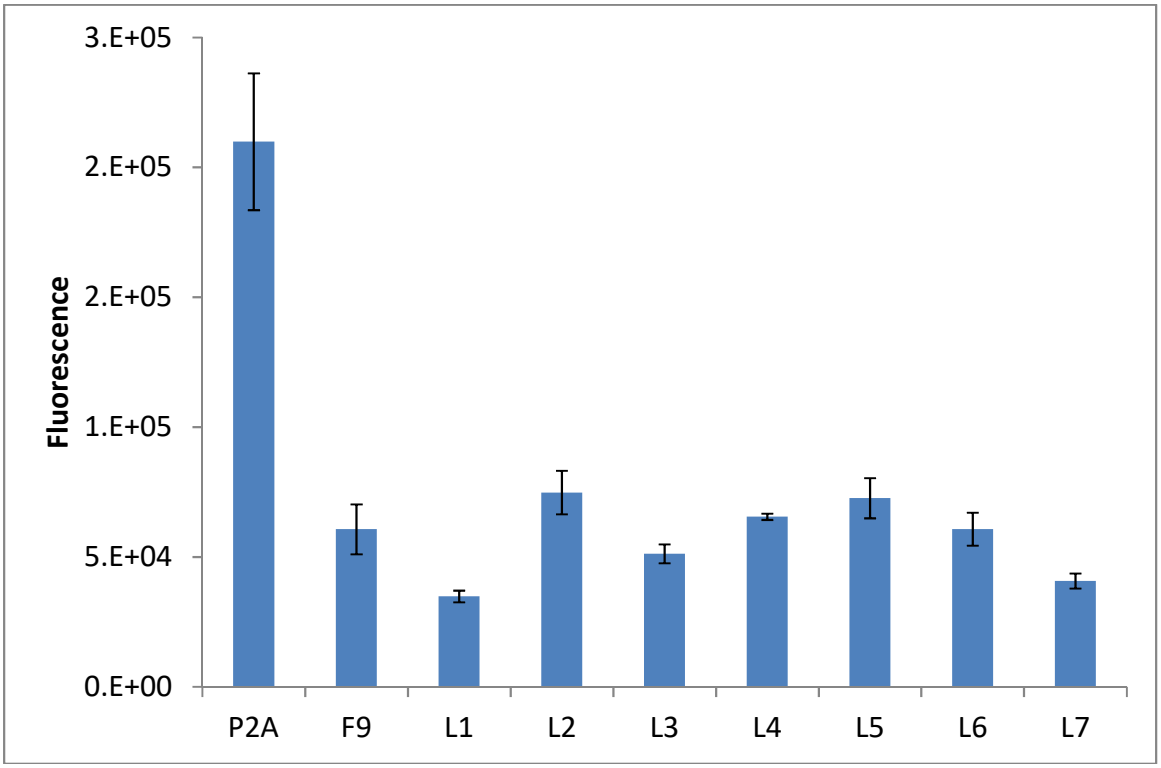
Figure 6-5 (continued): Leftmost of each is a control sequence (either P2A or T2A), with a set of sequences isolated from individual library sorts and re-transformed to fresh strains. **A**, **B** and **C** represent sequences isolated from three separate independent libraries, each individually screened by FACS.

In a final approach to identify novel high-activity 2A sequences, several of the most active variants identified above (F6, F9, J3 and J6) were mutagenized by error-prone PCR to improve this activity. As discussed above, it was hypothesized that the sequence space is much too large to identify optimal sequences using this library approach. However, it is possible that one of these novel, minimally active sequences could be significantly improved to the level of native 2As or better by only 1 or 2 mutations, yet still be sequentially distinct from any of the native sequences.

The four sequences to be tested were pooled and a single library was made using an error-prone PCR insert to re-mutagenize the variable region of the 2A sequence. The library was again made by Gibson assembly, with ~1.12 x 10^5 unique variants created.

This library was again sorted using FACS to isolate the brightest 0.1% of cells. Next, plasmids were isolated and sequenced, confirmed to each contain 1-2 coding mutations from one of the parent sequences, and unique variants were re-transformed to fresh strains and tested. Unfortunately, none of these exhibited any improved activity (Fig 6-6).

Figure 6-6: 2A activity of error-prone library sort isolates, as measured by cell fluorescence from the TEF-Sec-2A-YFP assay system

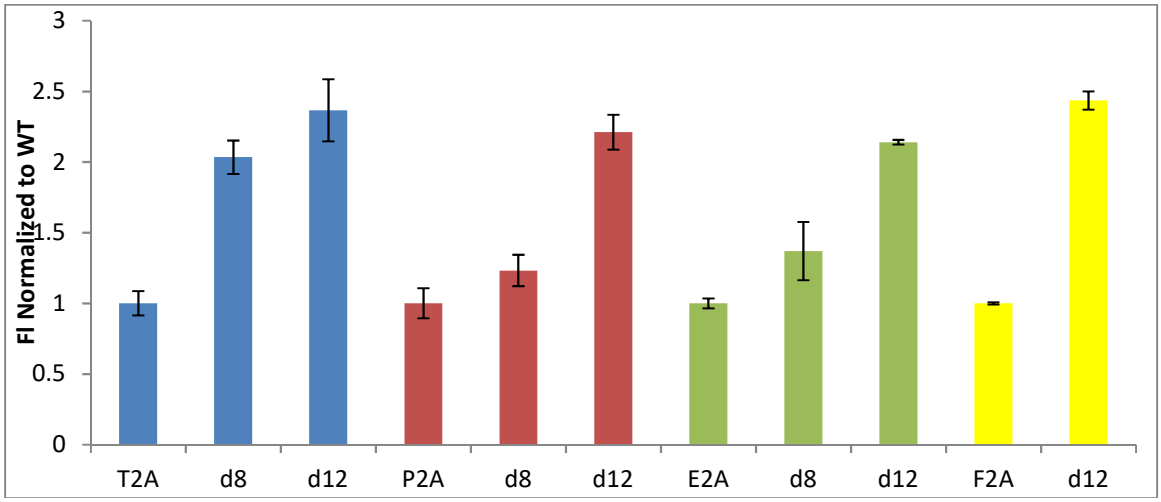


Leftmost is a control using P2A, and the next is a control "F9", one of the minimally active novel 2A sequences used to generate an error-prone library. Remaining are a set of sequences isolated from sorted error-prone library generated from these minimally active novel 2A sequences.

6.3.3 Characterizing Novel 2A Spacer Sequences

As discussed in section 6.3.2, two novel sequences were isolated from a library sort that exhibited significantly increased cleavage activity, and consisted of a complete T2A sequence with an upstream "spacer" sequence appended upstream. While these spacer elements did not exhibit any cleavage on their own (Fig 6-4b), through some unknown mechanism they were able to increase the cleavage activity of a downstream 2A site. To characterize this effect, both sequences were next cloned into the Sec-2A-YFP assay system, appended to the other three native 2A sites tested (P2A, E2A, and F2A). In all cases, appending either spacer element significantly improved cleavage activity by this assay (Fig 6-7).

Figure 6-7: 2A activity of Spacer-appended native 2A constructs, as measured by cell fluorescence from the TEF-Sec-2A-YFP assay system



The four native 2A sequences are tested, along with constructs containing the d8 and d12 spacers appended upstream. For each set, the fluorescence value is normalized to that of the native sequence. In each case, the spacer significantly increases activity.

These results were very encouraging, as they present a potential option for the use of 2A sequences for high-efficiency pathway expression. Each combination of spacer and native can exhibit relatively high efficiency, approaching 100%. But by mixing and matching spacers with different native 2As, sequential diversity can be established for multiple sites.

6.3.4 Utilizing 2A Sequences for Pathway Engineering Applications

Next, in order to demonstrate the utility of the novel, high-efficiency 2A spacer elements described above, several pathway engineering applications were tested. In each, a pathway of interest to ongoing metabolic engineering efforts was chosen, and the high-efficiency 2A sequences were used to demonstrate their utility compared to multiple vector approaches.

6.3.4.1 Beta-Carotene Production

One successful application of 2A sites to the metabolic engineering of *Saccharomyces cerevisiae* is in the production of beta-carotenes [74]. In this effort, genes for three heterologous pathway enzymes (*crtI*, *crtE* and *crtYB*) were assembled using T2A linkers, as described above (6.3.1). There is strong interest in producing high titers of various food-grade beta-carotene compounds in yeast, which can then serve as pigments or additives in the food industry [155]. This effort resulted in the successful production of beta-carotenes, although titers were not compared to the expression of the genes on separate vectors.

As a demonstration of the utility of the novel 2A spacer sequences, this pathway was reconstituted with several variations. First, the same T2A variants used previously

was constructed as a control. Next, a negative control was constructed using P2Ad, the non-functional mutant variant that does not allow cleavage [154]. Finally, several combinations of d8- and d12-spacer linked 2As were inserted into the same pathway. All constructs used the same genes and ordering as demonstrated previously.

Initial tests revealed surprising results: colonies from each individual transformation all exhibited similar levels of pigment, including the negative control. That is, the same 2A-linked beta-carotene pathway reported to result in activity in previous experiments *also* results in similar activity when constructed without polypeptide cleavage. These results suggest that the three-enzyme polypeptide may simply function as a fusion protein when expressed with a ~20-amino acid linker. Unfortunately, such a control did not seem to be included in the previous experiments [74], so it is difficult to determine how much of the beta-carotene production reported could be achieved without the 2A sites included. Further experiments were discontinued; if such a pathway functions without cleavage, it would not truly demonstrate the utility of high-activity 2A sequences.

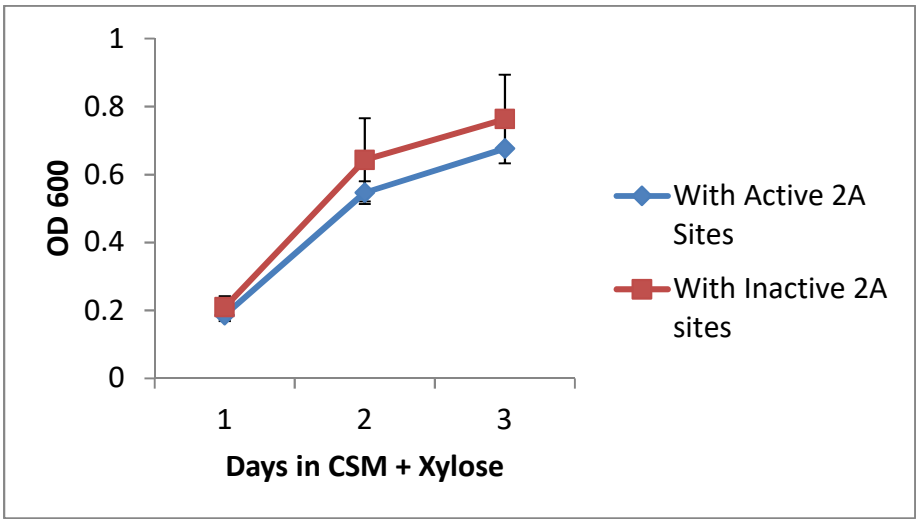
6.3.4.2 Xylose Consumption

A second pathway of interest is that of xylose catabolism, which has been studied extensively as an industrially relevant and desirable phenotype to develop in *Saccharomyces* [109]. A functional pathway for the efficient catabolism that has been previously demonstrated in yeast relies on the xylose isomerase gene *xylA* from *Piromyces spp.* [17] combined with the overexpression of native xylulokinase *xks1* [110]. Since co-overexpression of these two enzymes is necessary for efficient activity, this pathway serves as an excellent candidate for the demonstration of the utility of high-

activity 2A sequences. In fact, a 2A-linked xylose catabolism pathway consisting of *xylA* and *xks1* was utilized in the ICE evolution of xylose catabolism in Chapter 2.

First, however, the relevance of 2A activity needed to be demonstrated independently, to verify that polypeptide cleavage was limiting. Two constructs were made, with the two catabolic genes linked by P2A (as in Chapter 2) or with the two linked by P2Ad, the non-functional variant. Both were transformed into yeast and tested by growth in xylose-containing media (Fig 6-8). Unfortunately, it appears that again, cleavage is not necessary for the activity of these two enzymes, and this could not serve as a test case for demonstrating 2A activity.

Figure 6-8: Growth rate of strains expressing *xylA* and *xks1* linked by either P2A (an active 2A site) or P2Ad (an inactive variant)

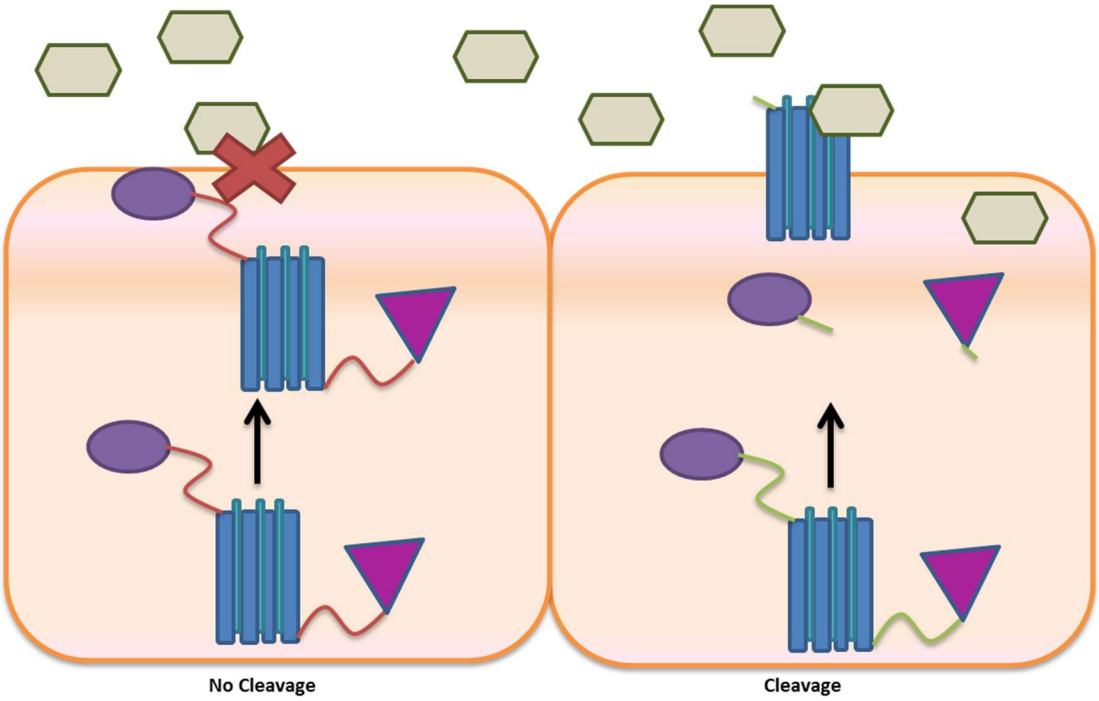


Growth in complete media containing xylose as the sole carbon source. Growth rates were similar in both, suggesting cleavage is not necessary for enzyme activity.

Next, several variations of the xylose catabolic pathway were created instead, focused on the expression of efficient xylose transports [156]. Because the transporter must be localized to the membrane for activity, while the downstream catabolic enzymes

were cytoplasmic, it was theorized that 2A cleavage must be necessary for activity (Fig 6-9).

Figure 6-9[4]: Schematic illustrating use of transporter membranes to demonstrate 2A activity through growth



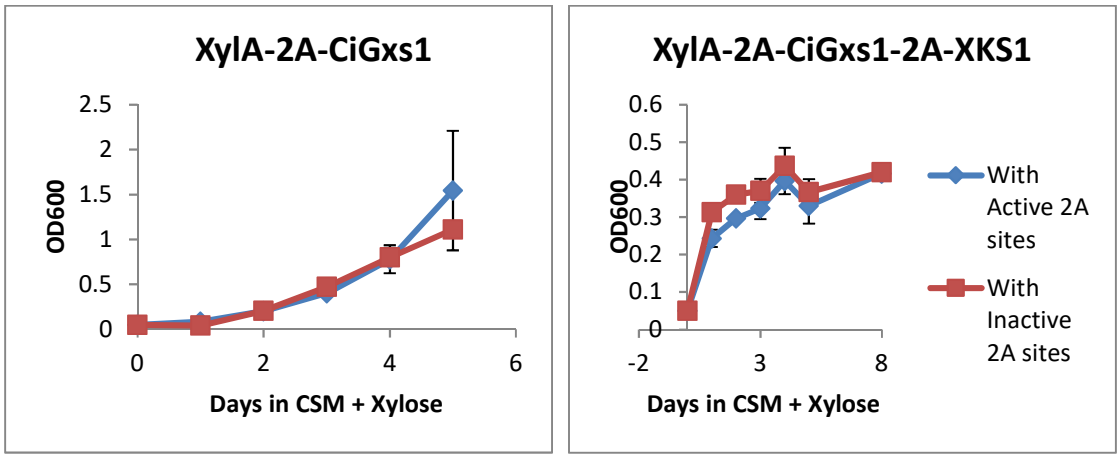
By linking the transporter protein to two other enzymes, it was theorized that 2A cleavage activity would be necessary for localization and thus growth.

Several constructs were made incorporating the xylose transporter in different combinations. The first consisted of *xylA* linked to an efficient xylose transporter CiGxs1-FIM [156], and a second included both *xylA* and *xks1* linked to the transporter, as shown in Fig 6-9. Two variants were made for each, using either native 2A sequences (P2A and T2A for the three-enzyme system) or the non-functional P2Ad sequences. All were transformed into yeast lacking all other hexose transporter [157] and tested by

[4] Schematic created by Jo Villa

growth in xylose-containing media (Fig 6-10). Unfortunately, it still appears that cleavage is not necessary for transporter activity in either case, and this would also not serve as an ideal test case for demonstrating 2A activity.

Figure 6-10: Growth rate of strains expressing *xylA* and/or *xks1* linked by either P2A and T2A (active 2A sites) or P2Ad (an inactive variant) to the xylose transporter *CiGxs1*,



Growth in complete media containing xylose as the sole carbon source. Growth rates were similar in both, suggesting that either cleavage is not necessary for enzyme activity, or that no activity is present even with cleavage.

6.3.4.3 Aromatic Amino Acid Biosynthesis

Another current goal relevant to metabolic engineering is the overexpression of aromatic amino acids and aromatic amino acid-derived compounds (see Chapter 5). As part of the evolutionary engineering efforts described in Chapter 5, several pathways of amino acid biosynthetic enzymes were constructed using 2A sequences. One consisted of three genes, *aro2*, *aro3*, and *aro4*, which encode chorismate synthase and two versions of DAHP synthase respectively. This construct was made as a poly-gene cassette using P2A

and T2A to link the three enzymes, in the order listed above, a construct called “AROp”. The goal was to co-evolve this cassette using ICE, in order to identify potentially synergistic mutations between the three genes.

However, when expressing this construct in yeast, a very unexpected result was observed: expressing the poly-gene 2A construct alone significantly *increased* aromatic amino acid production (Fig 6-11). *aro3* and *aro4*, both encoding DAHP synthase variants, are known to be feedback inhibited by phenylalanine and tyrosine, respectively; furthermore, it has been shown that several *aro4* mutations can remove this feedback inhibition [147], leading to significantly elevated production of aromatic amino acids when expressed [148]. Through some unknown mechanism, connecting these three genes using 2A sites also increased their activity, potentially by disrupting this feedback inhibition domain.

Figure 6-11: Tyrosine production measured from several strains containing aromatic biosynthetic genes

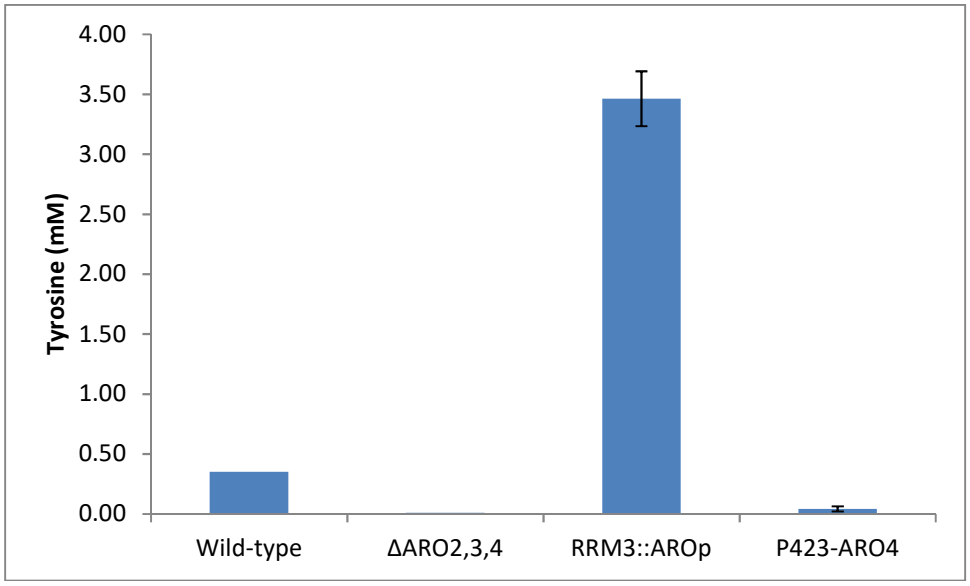
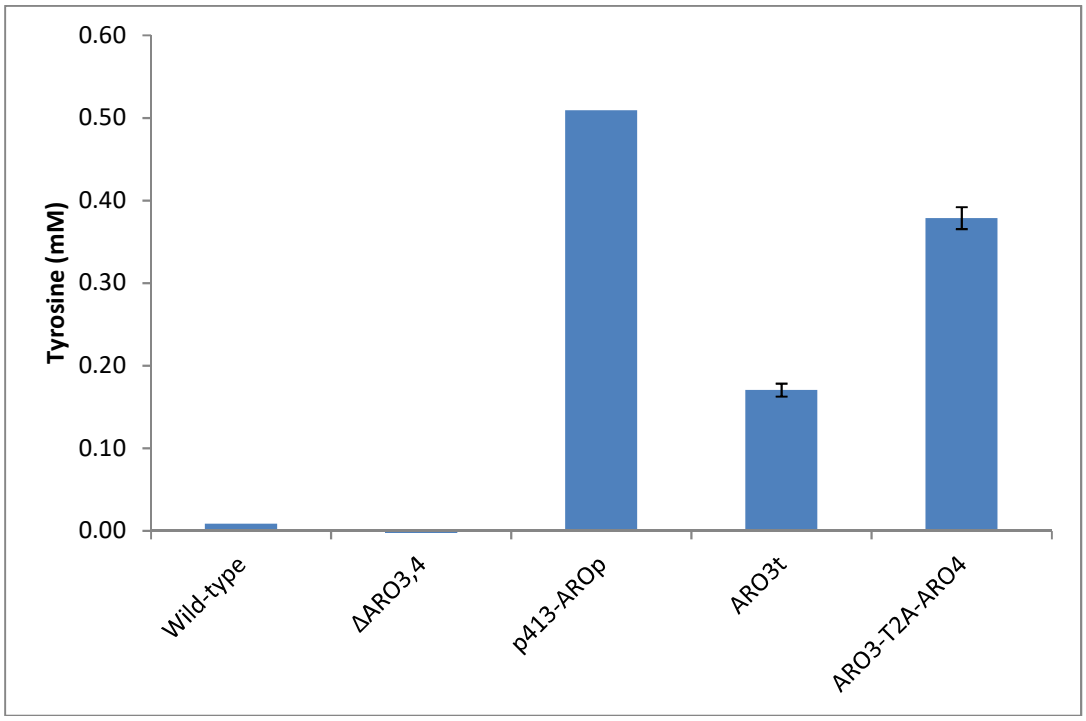


Figure 6-11 (continued): From left: wild-type By4741; By4741 with the three biosynthetic genes *aro2, aro3,* and *aro4* knocked out; the ARO2-P2A-ARO3-T2A-ARO4 poly-gene cassette integrated into the *rrm3* locus in the triple knockout strain; the wild-type *aro4* expressed on a high-copy plasmid in the triple knockout strain. The 2A-linked poly-gene cassette significantly boosts production over wild-type, while strongly overexpressing wild-type *aro4* does not increase production even to wild-type levels.

To explore this further, several variations of this gene cassette were constructed. One was made by simplifying expression to only *aro3* and *aro4*, linked by a single T2A site. A second construct, denoted "ARO3t", was made by expressing the post-cleavage sequence of the ARO3 protein that would result upon successful 2A cleavage from the AROp polypeptide. That is, a single proline reside was added to the N-terminus, while a 20-amino acid long tail was added to the C-terminus (the remaining sequence of T2A post-cleavage). Both of these were cloned into a low-copy plasmid, and transformed into a double knockout strain lacking *aro3* and *aro4*, and thus with no DAHP synthase activity; tyrosine production was again quantified (Fig 6-12).

Figure 6-12: Tyrosine production measured from several strains containing AROp variants



Wild-type By4741; By4741 with the two biosynthetic genes aro3, and aro4 knocked out; the ARO2-P2A-ARO3-T2A-ARO4 poly-gene cassette on a low-copy plasmid in the triple knockout strain; ARO3t on a low-copy plasmid in the double knockout strain, and ARO3-T2A-ARO4 on a low-copy plasmid in the double knockout strain.

Interestingly, there is no single clear cause for the significant overproduction afforded by the AROp system. ARO3t, the modified ARO3 protein that would result from cleavage of AROp, boosts tyrosine production significantly, suggesting that the C-terminal tail may have some effects on reducing feedback inhibition. Including *aro4* with a T2A linker boosts production still further, suggesting that it, too, may have some effect. However, cells expressing the full AROp system remains the highest producer observed,

implying there may be some additional effect from also incorporating the ARO2 protein as well.

While these data certainly merit further investigation for protein engineering and aromatic production in yeast, they also imply that such a system would not be useful for demonstrating the utility of 2A sites. Unfortunately, the very act of linking the proteins in a 2A-linked poly-gene cassette vastly impacts protein activity. If such a cassette were used in a directed evolution experiment, it is unlikely that any beneficial mutations could be identified that would also apply to the unlinked proteins. Further, in this case protein activity seems not to be limited by 2A activity, so employing the novel high-activity variants in this system would not illustrate their potential impact.

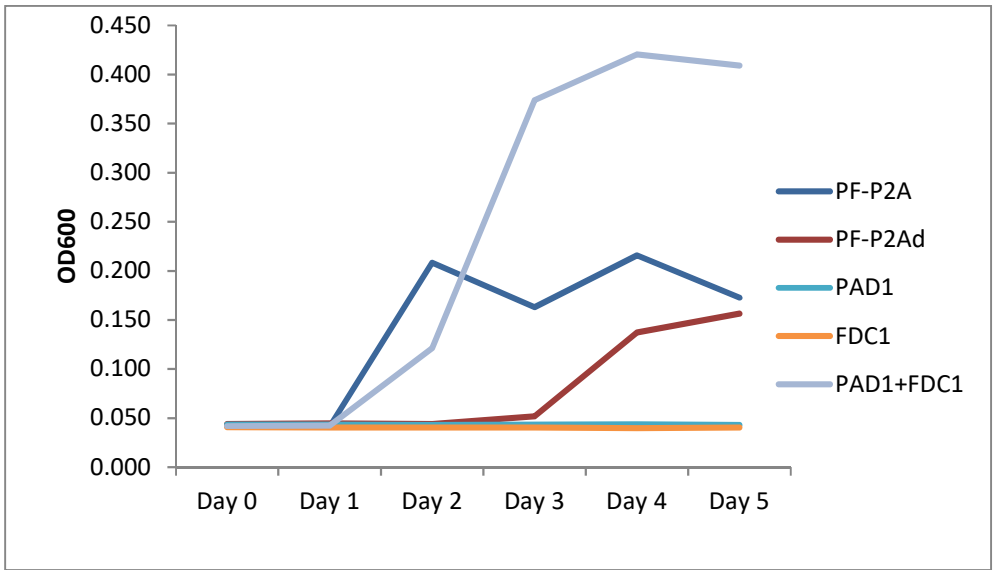
6.3.4.4 Cinnamic Acid Decarboxylation

The final system in which high-activity 2A sites were employed to demonstrate their utility is in the decarboxylation of cinnamic acid. Engineering more efficient decarboxylation is of high interest in several metabolic engineering efforts; for example, decarboxylation of PCA remains the limiting step in muconic acid production [4, 158], and decarboxylation of cinnamic acid can be used to produce styrene [159]. Cinnamic acid decarboxylation was chosen as a demonstration of 2A utility because a recent experiment found that co-overexpression of two native genes, *pad1* and *fdc1,* significantly improved cinnamic acid decarboxylation, while expression of each individually did not [159]. Furthermore, these two genes are not co-localized; PAD1 is transported to the mitochondria while FDC1 acts in the cytoplasm. Interestingly, it has recently been reported that PAD1 may not directly exhibit decarboxylation activity, but instead catalyze the production of a limiting co-factor [160]. Since activity requires

overexpression of both genes, and since a fusion protein would be non-functional due to co-localization, this system serves as an ideal test case for demonstrating potential improvement high-activity 2A sites can afford.

First, the same two control constructs were made to ensure such a system would function as expected: *pad1* and *fdc1* connected with either the functional P2A or the non-functional P2Ad. Since cinnamic acid inhibits cell growth, these constructs were transformed into yeast and then simply grown in the presence of cinnamic acid to compare activity (Fig 6-13). These initial controls demonstrate that growth in cinnamic acid is a good measure of activity, since over-expression of each individual gene did not allow any growth, while co-overexpression of both simultaneously did. Second, the strain expressing *pad1-P2A-fdc1* gene *did* exhibit growth, although not as much as the two genes separately expressed. Finally, the same construct with P2Ad exhibited only minimal growth, confirming that cleavage is necessary for activity in this system.

Figure 6-13: Growth rate of yeast strains expressing various combinations of *pad1* and *fdc1* in cinnamic acid-containing media



Growth in complete media containing an added 0.65 mM cinnamic acid. "PF" strains contain *pad1* and *fdc1* connected with either P2A or P2Ad site; the other three contain the same genes expressed individually or concurrently, but each under the control of an individual promoter (TEF in all cases) and terminator.

Next, variations of this cassette were constructed to compare and contrast the native 2A sequences with the d8 and d12 spacer elements. This set was assayed in a more quantitative fashion; each strain expressing the variation was incubated in a solution with 1 mM cinnamic acid, and timepoints were taken over 4 days. Each timepoint was then assayed for cinnamic acid concentration using HPLC.

First, the controls performed as expected: strains harboring an empty vector, overexpressing either gene individually, or the two linked by non-functional P2Ad, did not result in any cinnamic acid degradation, while a strain expressing both concurrently resulted in complete degradation (Fig 6-14a). Next, the four native 2A sequences exhibited degradation of cinnamic acid commensurate with their relative activities as measured by the secreted fluorescent protein assay; that is, P2A and T2A enabled more

rapid consumption than E2A and F2A (Fig 6-14a). Finally, several variations of improved spacer sequences were tested, including d8-P2A, d12-P2A, and d12-T2A (Fig 6-14b). Unexpectedly, these performed worse than the native P2A or T2A sequences.

Figure 6-14: Cinnamic acid degradation after incubation with cells expressing various combinations of *pad1* and *fdc1*

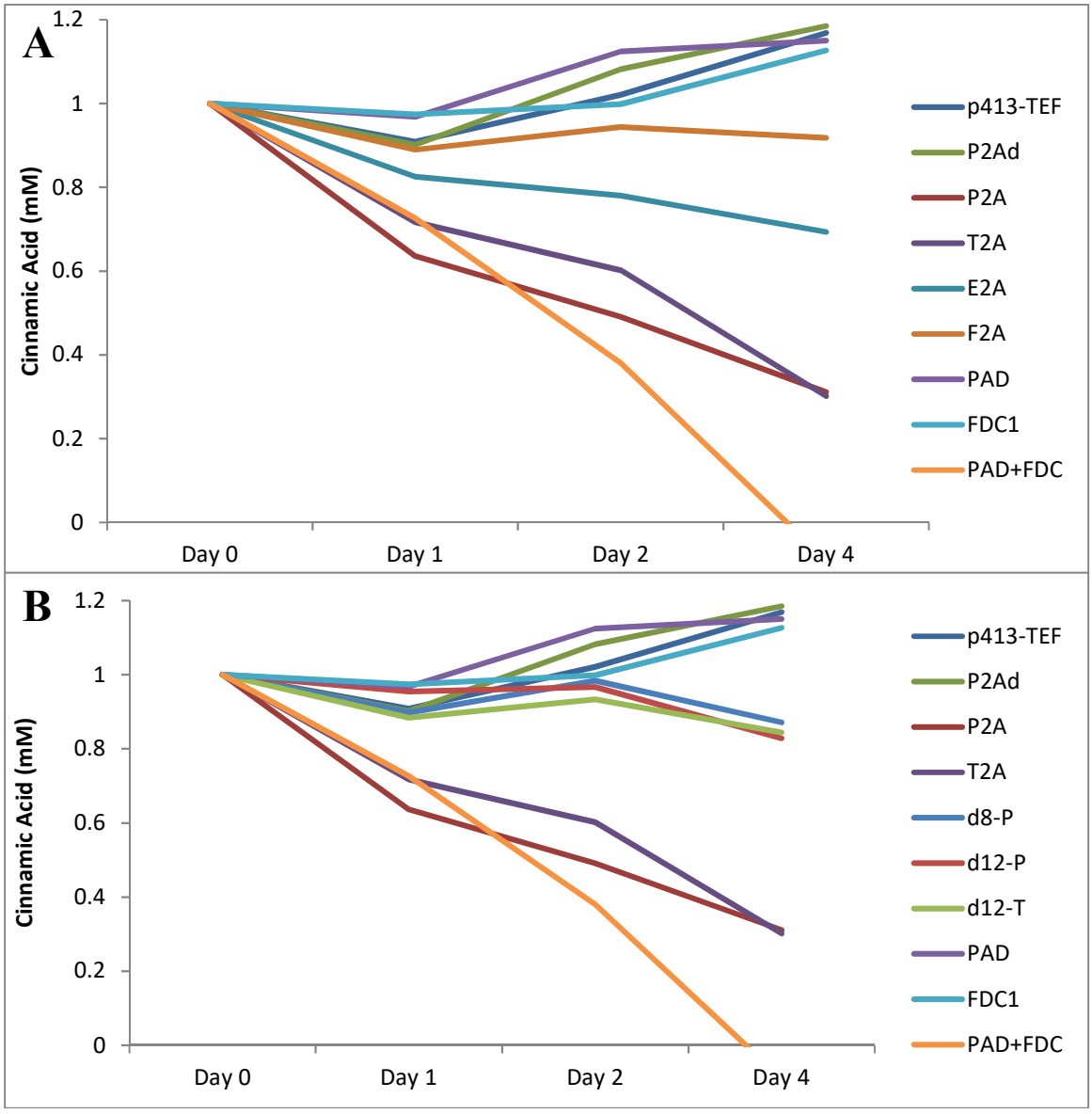


Figure 6-14 (continued): A) p413-TEF is the empty vector; P2Ad is the non-functional 2A variant; the next four represent the native 2A sequences; PAD1 and FDC1 each express a single gene, while "PAD1+FDC1" express both separately. **B**) The controls are the same as in the previous plot, but also including d8- and d12- spacer sequences, inserted upstream of either P2A ("-P") or T2A ("-T").

It is not clear why the spacer elements did not improve activity in this system, where it is seemingly limited by 2A cleavage efficiency. While there are several reasonable hypotheses, the fact remains that this system also does not support the idea that high-efficiency 2A elements would be useful for this pathway engineering effort. With this, further experiments to utilize 2A sequences with the cinnamic acid decarboxylation pathway were discontinued.

6.4 Discussion and Conclusions

Here, we characterize and investigate viral 2A sequences for their value in pathway expression and evolution. The secretion tag-based assay allows a novel method for quantifying 2A activity and allows us to further study which factors affect 2A cleavage in a way not previously possible. For example, the assay reveals that codon usage can significantly affect activity, something that was assumed to be neutral in prior demonstrations [74]. The screen also enabled the discovery of novel spacer elements than can significantly improve 2A activity. These sequences hold great promise for the expression and evolution of entire metabolic pathways simultaneously.

However, these data did not support the hypothesis that the spacer elements, which seemed to enhance 2A activity by the secretion assay, would always be preferable for use in pathway engineering efforts. In fact, in several applications tested, including

xylose catabolic pathways, aromatic amino acid biosynthetic pathways, and cinnamic acid decarboxylase pathways, the 2A sequences did not actually allow efficient poly-enzyme expression. In several cases this was due to simply to the superfluous nature of the cleavage; if uncleaved fusion proteins retain activity, then 2A cleavage is not limiting and the spacer elements are unnecessary for efficient expression. In several other cases, however, the use of 2A sequences impacted the enzyme activity. This is presumably due to the remaining polypeptide tag remaining post-cleavage, which can potentially disrupt folding or activity. While this was not always negative, such as in aromatic amino acid production, any such effect will interfere with evolutionary efforts, since beneficial mutations may or may not have the same effect when expressed as a single protein.

While we do not successfully demonstrate the utility of 2A sequences here, the challenges remain context-specific. When expressing metabolic pathways, 2A sites remain a viable option for concurrent expression and regulatory control. It is important to understand these limitations, however, and ensure that the 2A cleavage does not interfere with pathway activity. Furthermore, the novel assay remains useful in quantifying 2A activity, and future work can explore using this to further understand 2A cleavage and improve these sites for more widespread use.

Chapter 7 - Conclusions and Future Work

Taken together, this work has significantly advanced the state of the art for rapid evolutionary engineering of *Saccharomyces cerevisiae*. Novel tools have been developed, validated, and expanded to aid and allow the creation of new and valuable strains of yeast. Furthermore, these tools can be applied to a wide range of phenotypes, including both better and more efficient catabolism as well as production of high-value biochemicals. There remains opportunity, however, to continue to develop and expand these tools for even greater utility.

In Chapter 2, *in vivo* Continuous Evolution (ICE) was demonstrated as a powerful new tool for mutagenesis and the creation of libraries in yeast. ICE allows the creation of larger and more diverse mutant libraries with only a fraction of the time and effort needed for traditional techniques. This system was also well-characterized, demonstrating that the optimized synthetic retroelement functions as expected. Its utility was also validated by using it to evolve three very different phenotypes: enzyme specificity, in the case of *ura3*; complex tolerance phenotypes, in the case of *spt15*; and catabolic efficiency, in the case of the *xylA-xks1* pathway. With these successes, the utility of ICE has been demonstrated, and it promises to be a useful tool in a wide range of directed evolution experiments and efforts. Since ICE is essentially phenotype-agnostic, future work could potentially focus on using it to evolve non-growth phenotypes, to demonstrate its range. In addition, using the extremely high library size enabled by a bioreactor, difficult-to-engineer pathways or genes can be evolved, to demonstrate its power.

In Chapter 3, further characterization and study of ICE demonstrated how the process can be altered or controlled. Certain factors, such as the artificial intron, were shown to be non-impactful on the operation of ICE. Other factors, such as media

conditions and overexpression of *trans* "Helper" elements, were shown to have no beneficial effects on the operation of ICE mutagenesis. Several efforts have potential to expand the ICE system but were not able to be incorporated successfully in initial experiments, including integrase engineering, reverse transcriptase engineering, and the import of T7 RNA polymerase. Finally, several factors, such as the efforts to use Cas9 to target cDNA reintegration, showed promise but were not fully developed. Future work should focus on these efforts to better understand how ICE is still limited, and which factors could possibly relieve these bottlenecks. Incorporating additional orthogonal layers of control, especially by using Cas9 or T7 RNAP, have the potential to greatly improve the ICE system, by either allowing targeted re-integration and thus multi-cycle mutagenesis or enabling significantly higher error rates during library creation.

In Chapter 4, we introduced and validated a novel and powerful method for the creation of adaptable biosensors using RNA. By expanding the Spinach system to include additional secondary aptamers and then optimizing their use *in vitro*, we show how powerful such molecules can be as a screen for metabolite production. Furthermore, by creating and refining a model-based approach to design new biosensors, we demonstrate how the extremely large array of aptamers already reported can be used and adapted to quickly create novel Spinach-based biosensors. The utility was demonstrated by creating two completely novel metabolite biosensors: Spinach-Tyr1 aptamers that can detect aromatic amino acids, and Spinach-DNT1 that can detect 2,4-dinitrotoluene. This again confirms the versatility of such a method. Further work will focus on expanding this model to create more complex Spinach-based biosensors. It could, for example, be used to incorporate two or more metabolite-binding aptamers into the Spinach scaffold, using ΔG modeling to disfavor proper-folding unless all metabolites are present. Such a

molecule could function as a synthetic circuit, and its parameters could be easily tuned by changing the sequence as the model predicts.

In Chapter 5, we combine one of the novel Spinach-based biosensors created in Chapter 4 with the technology of microfluidic droplets to demonstrate how such biosensors can be used to rapidly evolve a secretion/production phenotype. By encapsulating yeast in droplets, it is possible to individually assay phenotype on a cell-by-cell basis at an extremely high rate of through-put. This enables screening for production phenotypes faster by orders of magnitude than traditional assay techniques. While this technology has previously been used to screen yeast libraries [79, 144], it has henceforth relied on very specific screening techniques such as exogenous enzyme-mediated oxidation. By importing and adapting Spinach-based biosensors to screen secretion/production directly, we demonstrate a much more powerful and generalizable screening technique, potentially applicable to a huge range of desirable phenotypes. We validated this method by evolving production and secretion of the aromatic amino acid tyrosine. For tyrosine, we mutagenized *aro4* and screened this library, identifying two novel mutations superior to any other reported variant. Previously, all *aro4* engineering has relied on crystal structure and knowledge of the protein dynamics [147]; this technique demonstrates for the first time an effective screen for the phenotype of interesting. Future work on this technology will focus on expanding it to evolve even more secretion or production phenotypes, to take advantage of all the Spinach-based biosensors characterized thus far. It will also focus on improving the screen efficiency, by optimizing the sorting device, fluorescence detection, and flow-rate, all of which could potentially increase the rate and efficiency of a sort by orders of magnitude [79, 145].

In Chapter 6, we investigated and characterized viral 2A elements, in the process discovering novel elements that significantly affect their activity. Such elements, which

allow co-translational polypeptide cleavage and thus the expression of multiple proteins from a single promoter, have great promise for engineering pathways. We developed a novel screen that can quickly and easily assay for 2A activity, which is critical when employing this sequences in a new pathway. Furthermore, a library screen of randomized 2A sites fortuitously discovered "spacer" elements, which can significantly enhance cleavage activity. With such elements coupled to relatively low-activity native 2A sites characterized thus far, it may be possible to express and evolve entire pathways using only a single genetic cassette. Unfortunately, we were not able to apply these novel 2A sequences in a metabolic engineering-relevant context. In some pathways, such as those used for xylose catabolism or beta-carotene production, cleavage efficiency was not a limiting factor; uncleaved 2A sites allowed functional fusion proteins. In others, such as cinnamic acid decarboxylase, the high-efficiency 2A sites actually decreased activity, potentially due to the significantly longer polypeptide sequence that remains on the C-terminal end of each cleaved polypeptide. Future work will continue to explore and characterize these sequences, in order to understand how and why the spacer elements function. Furthermore, additional pathways of interest could be explored for which a 2A site, or sites, enable high activity in a single cassette.

In summary, the work presented here represents a significant advancement in developing novel engineering tools for the evolutionary engineering of enzymatic pathways in *Saccharomyces cerevisiae*. These tools are expected to be applicable in a wide range of desirable uses, and can be employed as-is for ongoing efforts of strain engineering. This work also opens the door to more refinement and advancement of these tools, however, and offers a variety of opportunities for further directed evolution efforts.

Chapter 8 – Materials and Methods

8.1 Common Materials and Methods

8.1.1 Strains and Media

Yeast expression vectors were propagated in *E. coli* DH10β. *E. coli* strains were routinely cultivated in LB medium (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} sodium chloride) (Teknova) at 37°C with 225 RPM orbital shaking. LB was supplemented with 100 µg mL^{-1} ampicillin (Sigma) when needed for plasmid maintenance and propagation. Yeast strains for which maintenance of auxotrophic markers was unnecessary were propagated in YPD (10 g L^{-1} yeast extract, 20 g L^{-1} peptone, 20 g L^{-1} glucose), YPG (10 g L^{-1} yeast extract, 20 g L^{-1} peptone, 20 g L^{-1} galactose), or YPX (10 g L^{-1} yeast extract, 20 g L^{-1} peptone, 20 g L^{-1} xylose). When required for plasmid maintenance, yeast strains were cultivated on a yeast synthetic complete (YSC) medium containing 6.7 g of Yeast Nitrogen Base (Difco) L^{-1}, 20 g glucose L^{-1} and a mixture of appropriate nucleotides and amino acids (CSM, MP Biomedicals, Solon, OH). For YSC medium containing galactose or xylose, glucose was omitted from the above recipe and replaced with 20 g L^{-1} galactose or xylose, respectively. All components were supplemented with 1.5% agar for solid media.

For *E. coli* transformations, 25 µL of electrocompetent *E. coli* DH10β [161] were mixed with 30 ng of ligated DNA and electroporated (2 mm Electroporation Cuvettes (Bioexpress) with Biorad Genepulser Xcell) at 2.5 kV. Transformants were rescued for one hour at 37 °C in 1 mL SOC Buffer (Cellgro), plated on LB agar, and incubated overnight. Single clones were amplified in 5 mL LB medium and incubated overnight at 37 °C. Plasmids were isolated (QIAprep Spin Miniprep Kit, Qiagen) and confirmed by Sanger sequencing.

For yeast transformations, 50 µL of chemically competent *S. cerevisiae* BY4741 were transformed with 1 µg of each appropriate purified plasmid according to established protocols44, plated on the appropriate medium, and incubated for 3 days at 30 °C. Single colonies were picked into 1mL of the appropriate medium and incubated at 30 °C. Plasmids were isolated from yeast using a Zymoprep Yeast Miniprep Kit II (Zymo Corporation) and transformed into E. coli for further amplification.

8.1.2 Cloning Methods

8.1.2.1 Ligation cloning procedures

PCR reactions were performed with Q5 Hot-Start DNA Polymerase (NEB) according to manufacturer specifications. Digestions were performed according to manufacturer's (NEB) instructions, with digestions close to the end of a linearized strand running overnight and digestions of circular strands running for 1 hour at 37 °C. PCR products and digestions were purified with a QIAquick PCR Purification Kit (Qiagen). Phosphatase reactions were performed with Antarctic Phosphatase (NEB) according to manufacturer's instructions and heat-inactivated for 15 min at 65 °C. Ligations (T4 DNA Ligase, Fermentas) were performed overnight at 22 °C followed by heat inactivation at 65 °C for 20 min.

8.1.2.2 Recombination cloning in yeast

1 ug of each PCR fragment was digested with DpnI and co-transformed into *S. cerevisiae* BY4741 according to the procedure described in [162]. This transformation mixture was then plated on the appropriate dropout medium and allowed to grow for 3 days at 30 °C. Yeast colonies from this plate were scraped and plasmids were extracted (Zymoprep Yeast Miniprep Kit, Zymo Research). This plasmid mixture was then

transformed into *E. coli* DH10β and plated. Individual colonies were then amplified in liquid culture and plasmids were extracted. Correctly assembled plasmids were confirmed through restriction digestion and sequencing.

8.1.2.3 Gibson assembly

Isothermal assembly of DpnI digested PCR fragments was performed using the procedure of Gibson *et al* [163].

8.1.2.4 Vector and strain construction

For all knockouts, a loxP-*KanMX*-loxP deletion cassette was constructed from plasmid PUG646. One kilobase of homologous sequence was amplified from the region 5' to the desired integration site, and ligated at the 5' end of the loxP-*KanMX*-loxP module. A second kilobase of homologous sequence amplified from the region 3' to the integration site was then ligated at the 3' end of the loxP-*KanMX*-loxP module. The whole gene disruption cassette was then amplified and transformed into *S. cerevisiae* BY4741 or CEN.PK2, using a standard lithium acetate transformation method [162] or a version optimized for CEN.PK246, respectively. Cells were then plated onto YPD plus G418 plates (200 μg mL^{-1} G418). After one day of growth, microcolonies were replicated onto new YPD plus G418 plates. The resulting colonies were amplified in 3 mL YPD+G418 and the genomic DNA was extracted using the Wizard Genomic DNA Purification kit (Promega). Correct knockouts were confirmed by PCR.

Confirmed single knockout strains were transformed with the Cre expression plasmid pSH4744. Cre recombinase was then induced by incubating cells in YPG medium for 24 h. Cells were subsequently streaked onto YPD and replica-plated onto YPD plus G418 to isolate colonies in which the G418 marker (*KanMX*) had been excised. The Cre expression plasmid in G418-sensitive colonies was removed by incubating cells

in YPD plus 5-FOA for 24 h, thus excising the URA3-marked plasmid and yielding a clean version of the knockout strain with a single loxP site in the chromosome. Sequential gene knockouts were performed with the same protocol using this clean (*KanMX*-less) strain as a template, yielding a double-knockout strain.

Genomic integrations into the *TRP1* locus were performed with the aid of Cas9p [88]. Briefly, the parent strain was transformed with a *URA3*-marked plasmid expressing *CAS9* from the *pGAL1* promoter and a guide RNA targeted to the *TRP1* locus. Transformants were then transformed using standard procedures with a PCR product containing the retroelement of interest as well as homology to *TRP1*. This transformation mixture was then plated on media containing galactose to induce *CAS9* expression. Successful knock-ins were confirmed with PCR and the *CAS9* plasmid was excised through culture in 5-FOA.

8.1.3 Flow Cytometry Analysis

Yeast colonies were picked in triplicate from glycerol stock, grown in the appropriate medium to mid-log phase, and analyzed (LSRFortessa Flow Cytometer, BD Biosciences. Excitation wavelength: 488 nm, Detection wavelength: 530 nm). Day-to-day variability was mitigated by analyzing all comparable transformants on the same day. An average fluorescence and standard deviation was calculated from the mean fluorescence values for the biological replicates. Flow cytometry data was analyzed using FlowJo software (FlowJo, LLC).

8.2 Materials and Methods for Chapter 2

8.2.1 Plasmid Construction

All plasmids used in this chapter were constructed through yeast recombination cloning of PCRs, amplified using primers listed in Appendix A.

8.2.2 Analysis of transposition efficiency

For the following, it should be noted that the synthetic Ty1 retrotransposon (whether integrated on a plasmid or on the genome) contained an adjacent *HIS3* gene used for plasmid maintenance and as a selectable marker for genome integration, respectively. Therefore, we cultured all strains in histidine dropout media to facilitate media equivalence between all tested strains. As described in Chapter 2, the synthetic Ty1 retrotransposon contained an intron-containing *URA3* gene in the reverse orientation to provide a URA3^{+} phenotype if and only if retrotransposition occurs. Therefore, we plated induced cultures on media lacking histidine and uracil to count the number of cells containing an intact *URA3* gene, and hence which had a parent that underwent retrotransposition.

For low cell density induction, three biological replicates of a yeast strain carrying the engineered retrotransposon of interest were used to inoculate 50 mL liquid cultures lacking histidine and containing galactose, thus inducing retroelement transcription. After 3 days of growth at 30 °C, cultures were plated on agar containing glucose and either lacking histidine or lacking both histidine and uracil and allowed to grow for 3 days at 30 °C. Colonies were counted manually or through automated software [164] and counts were used as inputs to the Fluctuation Analysis Calculator [165] implementing the Ma, Sandri, and Sarkar Maximum Likelihood Estimation method [166]. Calculated mutation rates per cell were divided by the time spent in galactose medium to determine the transposition rate per cell per generation as well as 95% confidence intervals. This value was then used to estimate a library size.

For high OD tests with *S. cerevisiae*, cells were first cultivated in 50 mL liquid cultures lacking histidine and containing glucose and then resuspended in 50 mL liquid cultures lacking histidine and containing galactose to an initial OD of 1. After 3 days of growth at 30 °C, cultures were plated on agar containing glucose and either lacking histidine or lacking both histidine and uracil and allowed to grow for 3 days at 30 °C. Colonies were counted manually or through automated software [164] and counts were averaged. This average was used as an estimate for the number of transpositions which occurred during the experiment, and this average was divided by the total number of cells which had been exposed to galactose to obtain a per-cell measure of transposition rate.

8.2.3 qPCR Analysis

For determination of retroelement mRNA and cDNA levels, yeast strains carrying the appropriate retroelement were grown to mid-log phase (OD=0.5) in 5 mL YSC containing either glucose or galactose and lacking the appropriate amino acids. Total RNA was extracted (Ribopure Yeast Kit, Life Technologies) from half of each culture and converted to cDNA (High Capacity cDNA Reverse Transcription Kit, Life Technologies). Total DNA was extracted (Wizard Genomic DNA Purification Kit, Promega) from the other half of the culture. qPCR was conducted using 10ng of either cDNA or total DNA (FastStart SYBR Green Master, Roche) using primers specific for an intronless *URA3* (URA3RTPCRF and URA3RTPCRR) and with *ALG9* as an internal standard (Alg9F and Alg9R).

8.2.4 Next-Generation Sequencing

Ten replicates from BY4741 *Δrrm3* plus pGALmTy1-*Ty1RT-URA3I-pTEF1* were cultivated in 50 mL liquid cultures lacking histidine and containing glucose. After 3 days of growth at 30 °C, 1 mL culture was removed and the plasmids were extracted using Zymoprep Yeast Plasmid Miniprep Kit II (Zymo Research). The rest of the culture was then resuspended in 50 mL liquid cultures lacking histidine and containing galactose to an initial OD of 1. After 3 days of growth at 30 °C, 1 mL culture was extracted to obtain plasmids, and 1 mL culture was plated on agar containing glucose and either lacking histidine or lacking both histidine and uracil and allowed to grow for 3 days at 30 °C. Colonies were counted manually or through automated software [164] and counts were averaged. This average was then used as an estimate for the number of transpositions which occurred during the experiment. Two sequencing primer pairs with different barcodes were used to amplify the ampicillin sequence region from fresh pGALmTy1-*Ty1RT-URA3I-pTEF1* plasmid and pGALmTy1-*Ty1RT-URA3I-pTEF1* plasmid extracted from yeast grown in glucose medium, and 20 barcoded primer pairs amplified the *URA3* sequence region from 20 minipreps of galactose cultures. The PCR products were purified and the concentrations were determined by nanodrop. A final concentration of 50 ng μL^{-1} sample was prepared by combining 22 PCR purified products, with a 5:2 molar basis of ampicillin amplicon to URA3 amplicon. This mixture was then sequenced using an Illumina Miseq in 2x250bp paired-end mode.

Paired-end reads were matched up and error-corrected using pandaseq [167] using stringent quality filtering (threshold=0.95). Matched pairs were then divided up based upon barcode sequence using sabre [168], allowing for single nucleotide mutations (since each barcode differed by at least 2 bp from all other barcodes) and barcodes were removed with the trimmingreads.pl script of the NGS QC toolkit [169]. After combining

reads originating from the same culture into the same file, alignment to the unmutated amplicon was performed using ssaha2 [170]. Custom python scripts were then used to extract the total number of mutations identified, their locations on the wild-type sequence, and their frequency per mutant. All data originating from *URA3* was then averaged. Data was calculated by subtracting the mutation rates of the yeast-derived *Amp* amplicon from those of the averaged *URA3* amplicon. 95% confidence intervals for mutation counts were computed using the method of the Clopper-Pearson Interval [171].

8.2.5 Analysis of ICE mutation rate using *dKanMX* reversion assay

Strains *S. cerevisiae* BY4741 *trp1*::Ty1-*Ty1RT-dKanMX-pTEF1* and *S. cerevisiae* BY4741 *trp1*::*dKanMX-pTEF1* were cultured in 2.0 mL YSC−His medium containing 2% glucose. These cells were then resuspended into 2.0 mL YSC−His containing 2% galactose at an optical density of 1.0, in triplicate, for 3 days at 22 °C to induce mutagenesis. 1.0 mL of each culture was then plated on YPD containing 400 mg L^{-1} G418. After 2 days, colonies were counted to establish the rate of reversion mutations. 39 colonies from *S. cerevisiae* BY4741 *trp1*::Ty1-*Ty1RT-dKanMX-pTEF1* were grown, and genomic DNA was extracted. From each colony, the *dKanMX* marker was PCR amplified and sequenced to confirm the reversion mutations present.

8.2.6 Evolutionary Strategies

8.2.6.1 URA3 evolutionary strategy

During the first round of *URA3* evolution, a 50 mL culture containing *S. cerevisiae* BY4741 Δ*rrm3 trp1*::Ty1-*Ty1RT-URA3I-pTEF1* was mutagenized for 3 days at 22 °C at stationary phase and then transferred to 500 mL of galactose-containing,

uracil-deficient media and allowed to grow. Each day the optical density of this culture attained a value greater than one, the culture was re-inoculated into fresh medium containing an increased concentration of 5-FOA (0, 0.1, 0.3, 0.5, and 1.0 g L^{-1}). After each subculture, genomic DNA was isolated from a sample of bulk culture, and mutants were recovered for sequencing.

For the second round of evolution, a 50 mL culture containing *S. cerevisiae* BY4741 Δ*rrm3 trp1*::Ty1-*Ty1RT-ura3(3-5)I-pTEF1* was mutagenized for 3 days at 22 °C at stationary phase and then transferred to 500 mL of galactose-containing, uracil-deficient media with 0.3 g L^{-1} 5-FOA and allowed to grow. Each day the optical density of this culture attained a value greater than one, the culture was re-inoculated into fresh medium containing an increased concentration of 5-FOA (0.3, 0.5, 0.7 and 1.0 g L^{-1}). After growth in 0.5 g L^{-1} 5-FOA, the culture was split and re-inoculated into either 0.7 or 1.0 g L^{-1} 5-FOA. After each subculture, genomic DNA was isolated from a sample of bulk culture, and mutants were recovered for sequencing.

8.2.6.2 SPT15 evolutionary strategy

During the first round of *SPT15* evolution, 50mL cultures of *S. cerevisiae* BY4741 Δ*rrm3 trp1*::Ty1-*Ty1RT-SPT15I-pTEF1* were precultured in glucose media, after which they were induced in YSC−His + 20 g L^{-1} galactose for 3 days. Mutated cells were then pelleted and incubated in glucose-containing media for ~2 hours to allow recovery and time for expression of *Spt15* mutants. Next, butanol was added to a final concentration of 2.5% and cells were incubated with shaking for 3 hours in sealed flasks at 30 °C. At this point, genomic DNA was isolated from a sample of bulk culture and the *SPT15* gene was recovered for sequencing via PCR.

In order to undertake a second round of ICE, cultures of *S. cerevisiae* BY4741 Δ*rrm3* trp1::*Ty1-Ty1RT-spt15-B6I-pTEF1* were induced in YSC−His+gal for 3 days, pelleted, and incubated in glucose-containing media for ~2 hours to allow recovery and time for expression of *spt15* mutants. Next, butanol was added to final concentration of 3.5% and cells were incubated with shaking for 3 hours in sealed flasks at 30 °C. At this point, genomic DNA was isolated from a sample of bulk culture, and the SPT15 gene was recovered for sequencing via PCR.

8.2.6.3 Xylose isomerase pathway evolutionary strategy

Strains *S. cerevisiae* BY4741Δ*r-g-x trp1*::Ty1-*Ty1RT-XylAI-P2A-XKS1-pTDH3** and *S. cerevisiae* BY4741Δ*r-g-x trp1*::Ty1-*Ty1RT-XylA3*I-P2A-XKS1-pTDH3** were cultured in 50 mL YSC−His medium containing 2% glucose. These cells were then resuspended into 50 mL YSC−His containing 2% galactose at an optical density of one for 3 days at 22 °C to induce mutagenesis. Several 1 µL aliquots of culture were then plated on YSC−His or YPX plates. After 3 days, large colonies were identified. Large colonies were picked into a 96-well plate for growth analysis and clones exhibiting strong growth were subject to genomic DNA extraction and PCR to sequence mutant pathways. In this way, 3 pathway mutants were isolated.

8.2.7 Mutant Isolation

Genomic DNA from a small (~1 mL) volume of culture was isolated using crude total yeast DNA. For *SPT15* and *URA3* selections, this DNA was pre-digested with AscI (which exists uniquely in our synthetic intron) to enable specific amplification of spliced copies of the cargo. In general, unspliced copies of the cargo also amplified due to incomplete digestion, and so spliced copies were isolated through gel extraction. Then,

this extract was cloned into an expression vector for sequencing. For xylose pathway screening, *pTDH3*-XylA(3*)* and *pTDH3*-XKS1* gene fragments were individually PCR-amplified from total DNA extractions of isolates and cloned into an expression vector for sequencing.

8.2.8 Methods of characterizing ICE mutants

8.2.8.1 URA3 mutant growth analysis

First round *URA3* mutants were tested by plating on solid media containing lacking uracil and containing 0.3 g L^{-1} 5-FOA. After growth for 3 days, colony size was measured using a BioRad Gel Dock XR+ (Bio-Rad Laboratories, Inc.) and automated image analysis software [164]. After the second round of selection, the maximum rates of exponential growth were characterized for first and second-round mutants using a Bioscreen C (Growth Curves USA). Briefly, selected strains were inoculated at OD=0.1 and OD measurements were taken every 15 minutes using continuous shaking for 3 days at 30 °C. Growth rates were calculated using a custom MATLAB script (available upon request).

8.2.8.2 SPT15 butanol tolerance testing

To evaluate the effects of the mutant *Spt15-B6* gene, a *pTEF1-Spt15-B6* cassette was cloned into a low-copy plasmid and transformed into a fresh strain of BY4741 *Δrrm3*. Two control strains were also made, containing a blank *pTEF1* plasmid and a *pTEF1-SPT15* wild-type plasmid. Each was grown up in media, and then resuspended in fresh media at an OD of approximately 0.05. The actual OD was measured and recorded, and then butanol was immediately added to the desired concentration. The culture tubes were sealed to prevent evaporation, and the cultures were incubated at 30 °C. At various

timepoints, the optical density was measured, and normalized against the original OD of each replicate culture.

For high-butanol testing, the same strains described above were precultured in glucose media, then diluted to an OD=0.05 and incubated overnight. The next day, they were resuspended at OD=1.0 in media containing either 3.5% or 3.7% butanol. A time-zero sample was immediately plated to determine the number of cells in the culture. At intervals, a small volume was removed and diluted by 100-fold in water, then plated. On growth, the colonies were counted to create an estimate of how many cells remained alive. Each replicate was normalized to its own time-zero colony count.

For determination of wild-type or mutant *SPT15* expression levels, cells expressing either *pTEF1-SPT15* or *pTEF1-Spt15-B6-1* were grown to early exponential phase and total RNA was extracted and converted to cDNA as described above for "qPCR Analysis". qRT-PCR was carried out using primers that spanned the *pTEF1* promoter and the *SPT15* gene (primers Spt15-qPCR2f and Spt15-qPCR2r), eliminating amplification of endogenous *SPT15*. *SPT15* cycle threshold numbers were compared to that of a reference gene (*ALG9*) as above.

8.2.8.3 Xylose pathway mutant growth analysis

Colonies identified from plate-based selection were streaked onto a fresh plate and inoculated in 0.5 mL YSC−His+glucose medium for a 96-deepwell plate growth test. The original strains without galactose induction were also inoculated in 0.5 mL YSC−His+glucose medium in a 96-deepwell plate as controls. All cultures were then transferred to 1 mL xylose medium for growth test at an initial OD of 0.01. Growth curves were monitored by withdrawing 120 μL culture into a 96-well plate at each time point to read OD600nm using a plate reader (Cytation 3, BioTek Instruments, Inc.).

Isolates with improved growth rate were picked from the restreaked plate to extract genomic DNA as above. Each component of the pathway was then amplified and purified or ligated into *p413-pTEF1* for sequencing. Mutant strains were then inoculated in 1 mL YPX at an initial OD of 0.01 for growth test on a Bioscreen (Bioscreen C, Growth Curves USA).

For xylose isomerase kinetics analysis, mutant or wild type strains were grown until mid-exponential growth phase (OD of 0.6 to 0.8) in selective media. Cells were collected by centrifuging at 3,000 × g for 5 min at 4 °C. Total protein was extracted using Y-PER Plus (dialyzable yeast protein extraction reagent, Thermo scientific), and protein concentration was determined using the Pierce BCA protein assay kit (Thermo Scientific). Xylose isomerase activity from these cell extracts was determined by measuring oxidation of NADH at 340 nm using a spectrophotometer57. Each 1 mL reaction contained 100 mM Tris-HCl (pH 7.5), 0.15 mM NADH, 10 mM MgCl2, 2 U sorbitol dehydrogenase, and the diluted protein samples. Kinetic parameters were determined within a range of 25 to 500 mM xylose. All enzyme assays were performed in biological triplicate.

8.2.9 Comparison of *in vivo* and *in vitro* mutagenesis

To compare the ICE-derived *ura3* library to one which could be obtained using traditional methods, error-prone PCR libraries of *ura3* were also constructed. *ura3* was amplified in an error-prone fashion using the GeneMorph II Mutagenesis kit according to manufacturer instructions. This was done in duplicate to achieve either 4.5 – 9.0 mutations kb^{-1} (“Medium”) or 9 – 16 mutations kb^{-1} (“High”). Both were ligated separately into a low-copy vector with the *pTEF1* promoter, to match the construct made

for the ICE library. The mutant plasmid libraries of approximately 10^5 in size were then transformed into *E.coli* DH10-β using electroporation, which were then harvested from petri dishes. The plasmid libraries were then purified and transformed into wild-type By4741 yeast as described above. After transformation, the yeast libraries were grown to stationary phase. Concurrently, a new ICE-derived library of *pTEF1-ura3* was created as described above, however with a reduced galactose-induction volume of just 1 mL. After 3 days of induction, the two error-prone libraries and the ICE library were each resuspended in uracil-deficient media containing 0.3 g L^{-1} 5-FOA to a final OD of 0.05 and allowed to grow. Each day the optical density of this culture attained a value greater than approximately 3.0, the culture was re-inoculated into fresh medium containing an increased concentration of 5-FOA (0.3, 0.5, 1.0, 1.5, 2.0 g L^{-1}). At each subculture, a sample was plated. For characterization of adaptation, 5 colonies from each plate were picked and grown to stationary phase. Their growth rates in media containing various concentrations of 5-FOA were characterized using a Bioscreen C (Growth Curves USA). Briefly, selected strains were inoculated at OD=0.1 and OD measurements were taken every 15 minutes using continuous shaking for 3 days at 30 °C. Growth rates were calculated using a custom MATLAB script (available upon request).

8.3 Materials and Methods for Chapter 3

8.3.1 Plasmid Construction

All HIV-RT constructs were made by altering a previously constructed HIV/Ty1 retroelement [152]. Changes in the primer binding site and protein cleavage site were made by inverse PCR, with new or altered sequence included on the 5' end of the oligosaccharide. All primers are listed in Appendix A.

All reverse transcriptase mutants were made by either inverse PCR, as described above, or using QuikChange II Site-Directed Mutagensis kit (Agilent) as directed, using primers lists in Appendix A.

Overexpression vectors were made by PCR amplifying the native gene from wild-type strains and cloning using restriction digestion. Terminators were likewise amplified from wild-type strains and also cloned in downstream of the 3' LTR using restriction digestion.

"Helper" and "Mini" elements were made by inverse PCR, resulting in removing sections of plasmids that would not be included (the 3'LTR and the integrase/RT respectively). These linearized fragments were then circularized by blunt-end ligation.

8.3.2 Other Analysis

Analysis of transposition efficiency, next-generation sequencing, and *dKanMX* reversion was all done as in Chapter 2; See Section 8.2 for additional details.

8.4 Materials and Methods for Chapter 4

8.4.1 Synthesis of RNA Aptamers

To construct all RNA sensors, synthetic double-stranded DNA sequences (gBlocks) were obtained from IDT, incorporating the Spinach sequence with the Tyrosine aptamer inserted into a stem loop segment as described previously [133], as well as a T7 promoter sequence allowing transcription of the appropriate sequence. These sequences are listed in Appendix A. DNA templates were separately amplified from each plasmid by use of T7p-Spinach2-Fwd and Spinach2r, which created small fragments including a T7 promoter sequence allowing in vitro transcription. RNA was produced and purified

using the Ampliscibe T7-flash in vitro transcription kit using the appropriate PCR as a template (Epicentre).

To measure sensor fluorescence, RNA was diluted to ~2-10 uM (varying between experiments) in 50 mM Tris-HCl, pH 7.5, 125 mM KCl, 5 mM MgCl2, and heat denatured at 95ºC for 3 minutes. It was then incubated at 37ºC for 1-2 hours to allow proper folding. Finally, DFHBI or DFHBI-1T (Lucerna) was added to a final concentration of 10 uM (again varying between experiments), and the metabolite of interest was added to its final concentration. Fluorescence was measured continuously after addition of the metabolite, with an excitation wavelength of 460 nm and an emission wavelength of 500 nm for DFHBI or an excitation wavelength of 482 nm and an emission wavelength of 505 nm for DFHBI-1T.

For tests with cells and cell supernatants, cells were precultured in complete synthetic medium for at least 2 days. The OD was measured, and each culture was centrifuged, washed with water, and finally diluted to the same OD with minimal media, which contains 20 g/L glucose, 20 mg/L methionine, 12 mg/L adenine, 20 mg/L uracil, 20 mg/L histidine, and 100 mg/L leucine. Cells were incubated at 30ºC in this media overnight, and then centrifuged to pellet. When testing cells or cells supernatants, fluorescence was measured for several hours at 30ºC and again after an overnight incubation, also at 30ºC.

8.5 Materials and Methods for Chapter 5

8.5.2 Synthesis of RNA Aptamers

This was done as in Chapter 4; see section 8.4 for details.

8.5.2 Construction of Error-Prone *aro4* Libraries

Two libraries were made for the evolution of aro4: one derived from the wild-type gene and one from the K229L mutant, which has been shown to exhibit lower feedback inhibition [148]. For the K229L-based library, the mutant gene was PCR amplified in an error-prone manner using GeneMorph II kit (Agilent) following the protocol for "medium" error-rate (4.5-9 errors per kbp). The amplified product was purified using gel extraction, then digested and cloned into a backbone plasmid as described above. The mutant plasmid library (~5 x 10^4 in size) was then transformed into *E.coli* DH10-β using electroporation, which were then harvested from petri dishes. The plasmid libraries were then purified and transformed into a strain of yeast with *aro3* and *aro4* (both of which encode DAHP synthase activity) knocked out [4]. After transformation, the yeast libraries were grown to stationary phase and glycerol stocked.

The ARO4-Reg library was made slightly differently. In this, an insert of ARO4 was target for mutagenesis consisting of 212 bp, centered around the region known to code for feedback inhibition activity (bp 576 – 787). This piece was amplified in an error-prone manner using nucleotide analogs [172], with 20 uM 8-oxo-dGTP and 20 uM dPTP to achieve an error-rate of roughly 5-10 per kbp, using ARO4-reg-f with ARO4-reg-r. This PCR was purified and re-amplified using Q5 Hotstart polymerase (New England Biolabs) to extend the PCR with more homology to ARO4, using ext-ARO4-reg-f with ext-ARO4-reg-r. Next, a backbone plasmid was made by inverse PCR of p413-TEF-ARO4 using ARO4-reg-BB-f and ARO4-reg-BB-r. The plasmid library was made by Gibson assembly of these two pieces. Again, rhe mutant plasmid libraries (~1.3 x 10^5 in size) were then transformed into *E.coli* DH10-β using electroporation, which were then harvested from petri dishes. The plasmid libraries were then purified and transformed

into By4741 ΔARO3ΔARO4 yeast as described above. After transformation, the yeast libraries were grown to stationary phase.

8.5.3 Screening libraries in microfluidic droplet sorting

Each library was individually grown to saturation from frozen stock in the appropriate YSC dropout media over 2-3 days. They were then diluted by mixing in 100mM Tris (pH 7.5), 200mM KCl, 10mM NaCl, 10mM MgCl2, 10uM RNA aptamer, 2uM DFHBI-1T (Lucerna, Inc), 1mg/ml of double stranded salmon sperm DNA (Thermofisher), to a final cell density of 0.3. At this concentration of cells and size of droplets (40um), this results in approximately 1 in 10 droplets containing a single cell and 1 in 100 containing two cells. This culture was then encapsulated in droplets and incubated at room temperature for 48 hours prior to sorting.

To detect and sort microfluidic droplets using a fluorescence signal, we made use of a custom built fluorimeter and microscope. As droplets flow through the microfluidic channel, they are exposed to a 473 nm laser to excite individual droplets and the emission is measured using a PMT with a 517 nm bandpass filter. Labview software with a user set threshold for sorting controlled a field programmable gate array card to apply a voltage to a high voltage amplifier (Teck) to dielectrophoretically move the droplet into the sorted outlet while unsorted droplets continue unperturbed. Gating was used to sort out the top ~0.1% of droplets detected. Sorted droplets were opened using perfloruooctanol and cultured in dropout media for several days, before being frozen for further assays.

8.5.4 Recovering high-producing variants post-sort

Post-sort, total DNA from post-sort cultures were extracted (Zymo) and re-transformed to *E.coli*. Simultaneously, the original library was grown from glycerol stock and the same DNA extraction and transformation were carried out. Next, individual clones were picked from each, followed by plasmids isolation and sequencing. Those plasmids encoding unique mutations were re-transformed into a fresh strain of yeast containing the same double knockout of *aro3* and *aro4*. This strategy eliminates any possible strain adaptation that may account for high-production.

8.5.5 Assaying Aromatic Amino Acid Production of Cell Cultures

Cells were precultured in YSC dropout media for at least 2 days. The OD was measured, and each culture was centrifuged, washed with water, and finally diluted to OD=3.0 with minimal media, which contains 20 g/L glucose, 20 mg/L methionine, 12 mg/L adenine, 20 mg/L uracil, 20 mg/L histidine, and 100 mg/L leucine. Cells were incubated at 30°C in this media for 2 additional days, and then centrifuged to pellet. The media supernatant was then measured using a tyrosine derivitization assay. 100 uL of supernatant was mixed with an equal volume of solution containing 0.05% (w/v) 1-nitroso-2-naphthol, 50% ethanol, 10% (v/v) nitric acid and 0.25 g/l (w/v) NaNO2. The reaction was catalyzed by incubating at 55°C for 45 minutes, then read by fluorescence using an excitation of 485nm and emission of 590nm. A standard curve of tyrosine was used to quantify. This procedure is based on an assay described previously [149].

8.5.6 Characterization of Exponential Growth Rates

Maximum rates of exponential growth were characterized for cells expressing mutant *aro4* genes using a Bioscreen C (Growth Curves USA). Briefly, selected strains were inoculated into the appropriate media, either complete (YSC-Histidine) or minimal (20 g/L glucose, 20 mg/L methionine, 12 mg/L adenine, 20 mg/L uracil and 100 mg/L leucine), at OD=0.1 and OD measurements were taken every 15 minutes using continuous shaking for 3 days at 30 °C. Growth rates were calculated using a custom MATLAB script (available upon request).

8.6 Materials and Methods for Chapter 6

8.6.1 Plasmid Construction

The Sec-2A-YFP assay plasmids were constructed in series. First, YFP was amplified and cloned into a low-copy vector with the TEF promoter using Gibson assembly. Next, the SecG sequence was inserted upstream of the YFP by inverse PCR: the entire plasmid was amplified, with the SecG sequence appended by inclusion on the primers. This fragment was then circularized using blunt-end ligation. All of the variations of this plasmid, including swapping the different 2A sites, were done by Gibson assembly. The plasmid libraries were made by amplifying a primer with degenerate nucelotides (rand2A) to result in overlap to a backbone PCR. For the first library, these two pieces were transformed to yeast directly for assembly through homologous recombination. For the next three attempts, the two pieces were assembled using Gibson Assembly, then transformed into *E.coli* DH10-β using electroporation, which were then harvested from petri dishes. The plasmid libraries were then purified and

transformed into wild-type By4741 yeast as described above. After transformation, the yeast libraries were grown to stationary phase.

The beta-carotene pathway described was assembled previously [152]. All variations were made by PCR amplifying the plasmid backbone and individual 2A sequences with overlap, then assembled using Gibson Assembly.

Of the other pathways explored, each construct was made similarly: native genes (*xks1, aro2, aro3, aro4, pad1, fdc1*) were amplified from the wild-type genome. The 2A sites were amplified with homology and the full plasmids were assembled with Gibson assembly.

All primers used are listed in Appendix A.

8.6.2 Aromatic Amino Acid Biosynthetic Pathway Engineering

For these experiments, each construct was expressed in the relevant knockout strain; those expressing *aro3* and *aro4* in the double-knockout strain [4], and those expressing *aro2, aro3,* and *aro4* in a triple-knockout strain. To make this triple knockout, the double-knockout strain was transformed with a cassette PCR-amplified containing the *URA3* gene and homology to the *aro2* locus. These strains were plated on uracil-deficient media, and clones were tested by PCR confirmation to confirm deletion.

8.6.3 Cinnamic Acid Quantification

Cinnamic acid was quantified using a ThermoScientific UltiMate 3000 HPLC system, and an Agilent Eclipse Plus C18 column (3.0 x 150mM, 3.5 uM pore size). The mobile phase was a mix of 1% acetic acid in acetonitrile and 1% acetic acid in water at a flow rate of 0.3 mL/min. Flow began with a 40:60 v/v ratio and ramped for 3 minutes to 55:45, followed by an additional 6 minutes of isocratic flow. The UV-Vis detector was

operated at 278 nm, with isothermal conditions at 25ºC. Cinnamic acid was diluted in water to make a standard curve.

Appendix A – Primers

Table A-1: Primers used in Chapter 2 (IDT)

Primer Name	Sequence
GalCas9F	AAAAGTATCAACAAAAAATTGTTAATATACCTCTATACT TTAACGTCAAGGAGAAAAAACCCCGGATATGGACAAGA AGTACTCCATTGG
CYC1tCas9R	AGGGCGTGAATGTAAGCGTGACATAACTAATTACATGAT CACACCTTCCTCTTCTTCTTG
Cas9CYC1tF	AGGGCTGACCCCAAGAAGAAGAGGAAGGTGTGATCATG TAATTAGTTATGTCACGCTTAC
gRNACYC1tR	AAGTGGCACCGAGTCGGTGGTGCTTTTTTTGTTTTTTATGT CTCGAGCGTCCCAAAACCT
gRNAF	TGTACGCATGTAACATTATACTGAAA
gRNAR	AAAACGACGGCCAGTGAG
PacIP416F	/5PHOS/GG-TTAATTAA-GGTACCCAATTCGCCCT
NotICYCtR	/5PHOS/GCGGCCGC-CGAGCGTCCCAAAACCT
NotIgRNAF	GG-GCGGCCGC-AGACATAAAAAACAAAAAAAGCACCA
PacIgRNAR	GG-TTAATTAA-CTTTGAAAAGATAATGTATGATTATGCTTTCA
SacIGalF	CCCCCC-GAGCTC-ACGGATTAGAAGCCGCC
SpeXbaIGalR	GG-ACTAGT-TCTAGA-GGTTTTTTCTCCTTGACGTTAAAG
GalpF1	CTCACTAAAGGGAACAAAAGCTGGAGCTCCTAGTACGGA TTAGAAGCCGC
Gal1pFixR2	TATACTAGAAGTTCTCCTCGAGGCGGTAGAGGAATAAGA AGTAATACAAACCGAAAATGT
GalpF2	CAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCC TCACTAAAGGGAACAAAAGCT
GalpR2	TTCCATTGTTGATAAAGGCTATAATATTAGGTATACAGA ATATACTAGAAGTTCTCCTCG
TyH3GenomeF	GCGCGGGCAAAGCCCAAAAG
TyH3GenomeR	TGCGCAAGCCCGGAATCGAA
TyH3PCRF1	TATCAACAATGGAATCCCAACAATTATCTCAACATTCAC CCAATTCTCATGGTAGCGCCT
TyH3PCRR1	GAACGGTTTCAATTGGAGAAATTGGAACAGCCTTCAAAG CTGCAATCAGGTGAATTCGTT
TyH3PCRF2	CTTCTAGTATATTCTGTATACCTAATATTATAGCCTTTATC AACAATGGAATCCCAACAA
TyH3PCRR2	ACTTTTGGACCATCCATACCTGGTTTCAACTTAACTGGAA CGGTTTCAATTGGAGAAATT
TyH3PCRF3	CTCGAGGAGAACTTCTAGTATATTCTGTATACCTAATATT ATAGCC

TyH3FlankR	ACTTTTGGACCATCCATACCT
HIVnoATGF	GCTTTGAAGGCTGTTCCA
HIVRThomR	ACCGATTATTTAAAGCTGCAGCCCAAGCTTATCGATTTAC AAGATCTTTCTAATACCGGC
URA3F	TTGTAAATCGATAAGCTTGGGCTGCAGCTTTAAATAATC GGTGTTAGTTTTGCTGGCCGC
URA3AIR	GCAGAAAAGCCTCCTTTAGTCCATATTAACATACCCGCG ATGAAGGTTACGATTGGTTGA
URA3R	AAAAAATGAGCAGGCAAGATAAACGAAGGCAAAGATGT CGAAAGCTACATATAAGGAACG
URA3AIF	CTCGAATTTTTACTAACAAATGGTATTATTTATAACAGCC GCCCATGTCTCTTTGAGCAA
AInoass2F	TCTAGAGGATCCCCGGGTACCGAGCTCGAATTTTTACTA ACAAATGGTATTATTTATAAC
AInoass3F	GTATGTTAATATGGACTAAAGGAGGCTTTTCTGCAGGTC GACTCTAGAGGATCCCCGGG
LTRF1	TGTACTAGAGGATCTATTACATTATGGGTGGTATGTTGGA ATAGAAATCAACTATCATCT
LTRR1	CGCGCGTAATACGACTCACTATAGGGCGAATTGGGTACC TGAGAAATGGGTGAATGTTGA
PPTNotIF	GCGGCCGCGATCTATTACATTATGGGTGGTATGT
LTRflankR	AGGGTTTTCCCAGTCACG
HISpromF	CTTTGCCTTCGTTTATCTTGCC
HISpromR	ACCACCCATAATGTAATAGATCGCGGCCGCCTCTAGTAC ACTCTATATTTTTTTATGCCT
P416F	GGTACCCAATTCGCCCT
P416R	GAGCTCCAGCTTTTGTTCC
Ty1RTHDF1	CCGTTCACCTTCAATCGATGCTTCTCCACCGGAAAATAAT TCATCGCACAATATTGTTCC
Ty1RTHDR1	AAAGCTGCAGCCCAAGCTTATCGATCTAATGAATCCATT TGTTAGTTAATAGTTTAAATG
Ty1RTHDF2	CGCAGATAAGTGACCAAGAGACTGAGAAAAGGATTATA CACCGTTCACCTTCAATCGATG
Ty1RTHDR2	CATATTTGAGAAGATGCGGCCAGCAAAACTAACACCGAT TATTTAAAGCTGCAGCCCAAG
RTmutF	TCAACAGTAAGAAAAGATCATTAGAAGA
RTmutR	GGAAGGGATGCTAAGGTAGAG
HistermR	CACCGATTATTTAAAGCTGCAGCCC
LTRTEFF2	TACCACCCATAATGTAATAGATCGCGGCCGC-ATAGCTTCAAAATGTTTCTACTCCTTTT
URA3TEFR2	GCACGTTCCTTATATGTAGCTTTCGACAT-AAACTTAGATTAGATTGCTATGCTTTCTTT
RT-URA3-BBf	ATAAGCTTGGGCTGCAGC

RT-URA3-BBr	CGATCTAATGAATCCATTTG
RT-eGFPf	CATTTAAACTATTAACTAACAAATGGATTCATTAG-TGCGTAAAGGAGAAGAACTTTTCAC
RT-eGFPr	CCGATTATTTAAAGCTGCAGCCCAAGCTTATCGAT-TTAAACTGCTGCAGCGTAG
RT-LacZf	CATTTAAACTATTAACTAACAAATGGATTCATTAG-TGACCATGATTACGGATTCAC
RT-LacZr	CCGATTATTTAAAGCTGCAGCCCAAGCTTATCGAT-TTATTTTTGACACCAGACCAACTGG
RT-CAN1f	CATTTAAACTATTAACTAACAAATGGATTCATTAG-TGACAAATTCAAAAGAAGACGC
RT-CAN1r	CCGATTATTTAAAGCTGCAGCCCAAGCTTATCGAT-CTATGCTACAACATTCCAAA
Lacz-URA3-BBf	TAAATCGATAAGCTTGGGCTGC
Lacz-URA3-BBr	TTGACACCAGACCAACTGG
LacZ-eGFPf	CATTACCAGTTGGTCTGGTGTCAAAAATAA-TGCGTAAAGGAGAAGAACTTTTCAC
LTRCYCF2	CTATTCCAACATACCACCCATAATGTAATAGATCGCGGC CGC-ATTTGGCGAGCGTTGGT
URA3CYCR2	GTAGCAGCACGTTCCTTATATGTAGCTTTCGACAT-TTAGTGTGTGTATTTGTGTTTGCG
LTRGPDF2	ACCACCCATAATGTAATAGATCGCGGCCGC-AGTTTATCATTATCAATACTCGCCATTTC
URA3GPDR2	GAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACAT-ATCCGTCGAAACTAAGTTCTGG
PPTNotIF	GCGGCCGC-GATCTATTACATTATGGGTGGTATGT
reGREsptR	TAACTAACAAATGGATTCATTAGATCGATAAGCTTGGGC TGCAGCTTTAAATAATCGGTG-TCACATTTTTCTAAATTCACTTAGCACAGG
reGREsptF	TGTTATAAATAATACCATTTGTTAGTAAAAATTCGGCGCG CCAAAAGCCTCCTTTAGTCCATATTAACATAC-TGGCCGATGAGGAACGTT
reGREtefF	GATGATAGTTGATTTCTATTCCAACATACCACCCATAATG TAATAGATCGCGGCCGC-ATAGCTTCAAAATGTTTCTACTCCTTTTTTACT
reGREtefR	CTAAAGGAGGCTTTTGGCGCGCCGAATTTTTACTAACAA ATGGTATTATTTATAACAG-TAAACTTAGATTAGATTGCTATGCTTTCTTTC
Intron*ura3(3-5)*F	ATATGGACTAAAGGAGGCTTTTGGCGCGCCGAATTTTTA CTAACAAATGGTATTATTTATAACAG-GCCTATAGGTTCCTTTGTTACTTCT
LTRTEFF2	TACCACCCATAATGTAATAGATCGCGGCCGC-ATAGCTTCAAAATGTTTCTACTCCTTTT
Ty1URA3F	CATTTAAACTATTAACTAACAAATGGATTCATTAGATCG

	ATAAGCTTGGGCTGCAGCTTTAAATAATCGGTG-TTAGTTTTGCTGGCCGCA
Intron*ura3(3-5)*R	ATAATACCATTTGTTAGTAAAAATTCGGCGCGCCAAAAGCCTCCTTTAGTCCATATTAACATAC-CTTTTGATGTTAGCAGAATTGTCATG
(m)GAXnewintronF	CATCCCTGCGTCTACAGC
(m)GAXnewintronR	TGTAAATCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTTTTAAAACA-CCAGAACTTAGTTTCGACGGAT
(m)GA3XnewintronXylA3for	TGGTAGCACCAGAACCTTGATAC
TART-FullMCSF	CC-TTAATTAACCGCGGGGATCCCCCGGGGCGGCCGCGATCTATTAC
TART-FullMCSR	CC-TTAATTAACCTGCAGGGAATTCCTAATGAATCCATTTGTTAGTTAATAGTTTAAATGT
NotITEFF	ATAAGAAT-GCGGCCGC-ATAGCTTCAAAATGTTTCTACTCCTT
XmaITEFR	TCCC-CCCGGG-AAACTTAGATTAGATTGCTATGCTTTCT
SacIIXylAR	C-CCGCGG-TTATTGATACATCGCGACAATAGCC
XmaIXylAF	TCCC-CCCGGG-ATGGCTAAAGAATATTTCCCTCAAA
PacIXKS1F	CC-TTAATTAA-ATGTTGTGTTCAGTAATTCAGAGAC
EcoRIXKSR	G-GAATTC-TTAGATGAGAGTCTTTTCCAGTTCG
P2AXKS1F	/5PHOS/CAAGATGGTGATGTTGAAGAAAATCCAGGACCA-TTGTGTTCAGTAATTCAGAGACAGA
P2AXylAR	/5PHOS/TTTTAATAAAGAAAAGTTGGTAGCACCAGAACC-TTGATACATCGCGACAATAGCC
XylAMCSintronF	GCGCCGAATTTTTACTAACAAATGGTATTATTTATAACAGT-CCCGGGAAACTTAGATTAGATTGC
XylAMCSintronR	GCCAAAAGCCTCCTTTAGTCCATATTAACATACTGGCTAAAGAATATTTCCCTCAAATTC
XmaIGPDR	TCCC-CCCGGG-ATCCGTCGAAACTAAGTTCTGG
NotImegaGPDF	ATAAGAAT-GCGGCCGC-ATAGCTTCAAAATGTTTCTACTCCTT
(m)GA3XnewintronRTfor	ACCAAGAAGAACATTGCTGATGTG
(m)GA3XnewintronP2Arev	TGAATATGGAAAGAAGAATGGAGAACCGAAACAAACCTCAGGAAAACAAGAACTATACGAGGCTATTA-TCGCGATGTATCAAGGTTCTGG
XmaIXylA3F	TCCC-CCCGGG-ATGGCTAAAGAATATTTCCCTCAAA
XylAGPDnointronF	TTTGAATTTGAGGGAAATATTCTTTAGCCATATCCGTCGAAACTAAGTTCTGGTGTTTTA
XylAGPDnointronR	TAAAACACCAGAACTTAGTTTCGACGGATATGGCTAAAGAATATTTCCCTCAAATTCAAA

GREnewF	TTTCACAGGTAGTTCTGGTCCATTGGTGAAAGTTTGCGGC TTGCAGAGCACAGAGGCCGCAGAATGT-GAACAAAAGCTGGAGCTCCTAGT
GREnewR	ACGAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTG TCTCCACACCTCCGCTTACATCAACACCA-CTTGGCCTCCTCTAGTACACT
XbaIXKS1	CCCCCC-TCTAGA-ATGTTGTGTTCAGTAATTCAGAGA
XKS1XhoI	CCCCCC-CTCGAG-TTAGATGAGAGTCTTTTCCAGTTC
GRE3KOfor	GTAATATAAATCGTAAAGGAAAATTGGAAATTTTTTAAA GATAGCTTCAAAATGTTTCTA
GRE3KOrevnew	TTGTTCATATCGTCGTTGAGTATGGATTTTACTGGCTGGA-TTAAGCAAGGATTTTCTTAA
MRE115'F	CCCCCC-CGTACG-TCTTTCCAACAAACCAAGCG
MRE115'R	CCCCCC-GTCGAC-AGTCGAGTTTTATCGGATCTGAGC
MRE113'F	CCCCCC-ACTAGT-TTGTACTTGATCCCTATATTATATTATATCCTATTTATAAC C
MRE113'R	CCCCCC-CCGCGG-AGTTCTATTGTGTGTCCAGGC
pUG6KOPCRnewF	CCAGCTGAAGCTTCGTACG
pUG6KOPCRnewR	CCGGCAGATCCGCGG
CAC35'F	CCCCCC-CGTACG-GTGGTTTGTTGTCTGTCTGG
CAC35'R	CCCCCC-GTCGAC-CTTTGAAACTAAATTTGTATATTGTTTGTCAGAA
CAC33'F	CCCCCC-ACTAGT-CCTAAACGTTCCTTGAAGCCA
CAC33'R	CCCCCC-CCGCGG-GTTCGGCTTTGGACATTTCG
BYHIR35'F	CCCCCC-CGTACG-ACTAGCAATGATTCCGTTTTACATTT
BYHIR35'R	CCCCCC-GTCGAC-AATAAGCTTTATCTAGAATCTGTGTTGAGG
BYHIR33'F	CCCCCC-ACTAGT-GATGACCATATTTTGGAAGAAGTGTG
BYHIR33'R	CCCCCC-CCGCGG-CAAATCTTTATCGTAATCAGATAATTTTTCCAA
CAC25'F	CCCCCC-CGTACG-AGAAAGGTCCTCAGATTGAGC
CAC25'R	CCCCCC-GTCGAC-TGTCCTGCCCCTTTGCT
CAC23'F	CCCCCC-ACTAGT-TTTTTAATATATTTAATGCGGTACATAAGAATGCC
CAC23'R	CCCCCC-CCGCGG-TCACGAGAGATGAGTCCACC
BefICE2F	ATGATTCACTGTCACTTAGTGAGC
AftACE2R	CTAAGAGTCTGTTTAGATCAACAGTCT
BefRRM1F	CATAGAACCGAGTGTAACACCA
AftRRM1R	AGGATTCTCCGAATAACCTCTAGC
APL25'F	CCCCCC-CGTACG-TATCCTGATGGAGCACTTCG

APL25'R	CCCCCC-GTCGAC-AGTTGAAACTGTTTTTTAAGTGCAGT
APL23'F	CCCCCC-ACTAGT-CTATAAACGTCCGTTGTAGTGAACT
APL23'R	CCCCCC-CCGCGG-CCTGACATCTTTGGACGTGG
IMT4p	CCG-GAGCTC-AACATCCAGT
IMT4t	GGC-GGTACC-TTATAGTCTATAGCTTAAAT
sup4p	CCG-GAGCTC-ACCATCTTGG
IMT4p	CCG-GAGCTC-AACATCCAGT
XbaITAL1for	CCG-TCTAGA-ATGTCCTCCAACTCCCTTGAAC
XhoITAL1rev	CCCG-CTCGAG-TTAGAATCTGGCTTCCAATTGTTCC
SacItefTAL1for	CCG-GAGCTC-ATAGCTTCAAAATGTTTCTACTC
SacITAL1cyc1rev	GGC-GAGCTC-CAAATTAAAGCCTTCGAGCGTCC
MCS-Fwd-SpeI	G-ACTAGT-ATGTCTAAAGGTGAAGAATTATTCACTGG
MCS-Rev-2	CCCCG-CTCGAG- TTATTTGTACAATTCATCCATACCATGGG
P2AYFPGlcF	ACAAGATGGTGATGTTGAAGAAAATCCAGGACCA- TCTAAAGGTGAAGAATTATTCACTGG
P2AstrawGlcR	TTTAATAAAGAAAAGTTGGTAGCACCAGAACC- CTTGTACAGCTCGTCCATGC
SpeImStrawberryF	CC-ACTAGT-ATGGTGAGCAAGGGCG
EcoRImStrawberryR	G-GAATTC-TTTATCGATCGATT- CTACTTGTACAGCTCGTCCAT
MCS-Fwd-SalI-2	TAACGC-GTCGAC- ATGTCTAAAGGTGAAGAATTATTCACTGG
mStraw-YFPf	GATCCCCCGGG-ATGGTGAGCAAGGGCGAG
mStraw-YFPr	ACTCGAGGAATTC- TTATTTGTACAATTCATCCATACCATGGG
His3AIgenomeflankF	ATCGATAAGCTTGGGCTGC
ARTR	TTAATTCTTAGTATTCCATGTGTCTCGT
HispromSPT15F	AAAAAATGAGCAGGCAAGATAAACGAAGGCAAAG- ATGGCCGATGAGGAACGT
HistermSPT15R	AAGCTTGGGCTGCAGCTTTAAATAATCGGTG- TCACATTTTTCTAAATTCACTTAGCACA
Spt15TEFR	AACTCCTTTAAACGTTCCTCATCGGCCAT- AAACTTAGATTAGATTGCTATGCTTTCTTT
SPT15intronR	GTTAGTAAAAATTCAAAAGCCTCCTTTAGTCCATATTAAC ATACTGGCCGATGAGGAACG
SPT15intronF	AAATGGTATTATTTATAACAGTAAACTTAGATTAGATTGC TATGCTTTCTT
IntronSiteF	CGCCGAATTTTTACTAACAAATGGTATTATTTATAACAGT
IntronSiteR	CGCCAAAAGCCTCCTTTAGTCCATATTAAC
TEFmutssacF	CCC-GAGCTC-ATAGCTTCAAAATGTTTCTACTCCTT

p416ura3EcoR1rev	CCG-GAATTC-TTAGTTTTGCTGGCCGCATC
XhoISPT15R	CCCCG-CTCGAG-TCACATTTTTCTAAATTCACTTAGCACA
BefCAC2F	CGTTTCTGAGAGGTAACTGAGG
CAC2BegR	GATTCTTGCTGTGTATTTGG
CAC2EndF	TGATTTAGCATGGTCTGAGG
AftCAC2R	TTGTTGCTGTTGGTCATTGG
BefCAC3F	ACAACCACTTCACCCAAACCC
CAC3BegR	TTGGGAAGATGTAAATGAGG
CAC3EndF	AAGAAGATGGGTTAGTCAAGC
AftCAC3R	TGGAAATGTTGTAGAGTGGAGG
BefAPL2F	TTCTCAACCATCCAAGTCGG
APL2BegR	CCAACTCAACGCATAAATCC
APL2EndF	TAACGATGATGTGCTATTGG
AftAPL2R	CTTGTTGATCTTTCTTCCCACC
BefMRE11F	ATTGATGGCTGATGACGTGG
MRE11BegR	TTGGTAGAGTGACTTCTTGG
MRE11EndF	CCAACGAGCAAACCCAAACG
AftMRE11R	TGTGTTTGAGGGCTCCTTGG
BefBYHIR3F	TCTACGCGGTCCATAATCTCC
HIR3BegR	ATCTAGCGTAGGAGAAGAATTGC
HIR3EndF	ATTTGACAGCGTTTGCTTGG
AftBYHIR3R	TTGACGCAAAGGAAATGTGG
BefGRE3F	ATGGGCGCATTACTACAAGAAG
AftGRE3R	CTGTTTGACGCACTGATGGGT
BegGRE3R	GACCTTCGGAGATGGCTTTC
EndGRE3F	AACCATCCAGGCAGTACCAC
URA3RTPCRF	ATTGTTAGCGGTTTGAAGCAGGCG
URA3RTPCRR	GAGCCCTTGCATGACAATTCTGCT
Alg9F	ATCGTGAAATTGCAGGCAGCTTGG
Alg9R	CATGGCAACGGCAGAAGGCAATAA
Spt15-qPCR2f	GCAATCTAATCTAAGTTTATGGCCGATGAGGA
Spt15-qPCR2r	GTTGCTGGTTTTGTACCATCTCGATTCTGG
NGSAmp1F	ACTGAT-GCCAACTTACTTCTGACAACG
NGSAmp1R	GCTACC-CCGCCTCCATCCAGTC
NGSAmp2F	CGTACG-GCCAACTTACTTCTGACAACG
NGSAmp2R	TGACAT-CCGCCTCCATCCAGTC
NGSnointron1F	GATACA-ATCGCGGCCGCCC

NGSnointron1R	GGAACT-AGAATGGGCAGACATTACGAATG
NGSnointron2F	AGTCAA-ATCGCGGCCGCCC
NGSnointron2R	TAACCG-AGAATGGGCAGACATTACGAATG
NGSnointron3F	AGCTTT-ATCGCGGCCGCCC
NGSnointron3R	TACAAG-AGAATGGGCAGACATTACGAATG
NGSnointron4F	GGCTAC-ATCGCGGCCGCCC
NGSnointron4R	AAGCTA-AGAATGGGCAGACATTACGAATG
NGSnointron5F	ATACGA-ATCGCGGCCGCCC
NGSnointron5R	CTGATC-AGAATGGGCAGACATTACGAATG
NGSnointron6F	TTACTG-ATCGCGGCCGCCC
NGSnointron6R	AGTTCC-AGAATGGGCAGACATTACGAATG
NGSnointron7F	ACTTGA-ATCGCGGCCGCCC
NGSnointron7R	GATCTG-AGAATGGGCAGACATTACGAATG
NGSnointron8F	ACATCT-ATCGCGGCCGCCC
NGSnointron8R	AATCGT-AGAATGGGCAGACATTACGAATG
NGSnointron9F	GCCAAT-ATCGCGGCCGCCC
NGSnointron9R	CACTGT-AGAATGGGCAGACATTACGAATG
NGSnointron10F	AGAATC-ATCGCGGCCGCCC
NGSnointron10R	GCCTAA-AGAATGGGCAGACATTACGAATG
NGSnointron11F	CTGCAG-ATCGCGGCCGCCC
NGSnointron11R	ACATCG-AGAATGGGCAGACATTACGAATG
NGSnointron12F	ATCACG-ATCGCGGCCGCCC
NGSnointron12R	CACGTA-AGAATGGGCAGACATTACGAATG
NGSnointron13F	TCACAT-ATCGCGGCCGCCC
NGSnointron13R	TATAGA-AGAATGGGCAGACATTACGAATG
NGSnointron14F	TGCAAA-ATCGCGGCCGCCC
NGSnointron14R	GTGCCA-AGAATGGGCAGACATTACGAATG
NGSnointron15F	TGTTAG-ATCGCGGCCGCCC
NGSnointron15R	ATAGAA-AGAATGGGCAGACATTACGAATG
NGSnointron16F	TCGAAG-ATCGCGGCCGCCC
NGSnointron16R	GAATGA-AGAATGGGCAGACATTACGAATG
NGSnointron17F	TACAGC-ATCGCGGCCGCCC
NGSnointron17R	TCTGAG-AGAATGGGCAGACATTACGAATG
NGSnointron18F	CTATAC-ATCGCGGCCGCCC
NGSnointron18R	AGCTAG-AGAATGGGCAGACATTACGAATG
NGSnointron19F	CGGAAT-ATCGCGGCCGCCC
NGSnointron19R	GCCATG-AGAATGGGCAGACATTACGAATG

NGSnointron20F	CACGAT-ATCGCGGCCGCCC
NGSnointron20R	GCTCAT-AGAATGGGCAGACATTACGAATG
SacI-HSXf	CC-GAGCTC-GGGTTCATTACAGCAGC
PstI-HSXr	CC-CTGCAG-TATATTTACCATCAACTCCGC
EcoRITkc1F	G-GAATTC-TTTGAAAGATGATACTCTTTATTTCTAGACAG
SbfITkc1R	GG-CCTGCAGG-TATATATATATAACTGTCTAGAAATAAAGAGTATCATCTT
EcoRITkc5F	G-GAATTC-TTTGAAAGATGATACTCTTTATTTATATATATATATATATATATATA
SbfITkc5R	GG-CCTGCAGG-TATATATATATATATATATATATATAAATAAAGAGTATCATCTTTCAAA
EcoRITkc8SbfIF	G-GAATTC-TTTGAAAAAATTTATTAAAAAAAAAATATATA-CCTGCAGG-CC
SbfITkc8EcoRIR	GG-CCTGCAGG-TATATATTTTTTTTTTAATAAATTTTTTCAAA-GAATTC-C
URA3AICYC1F	CCG-GAATTC-CGAGCGTCCCAAAACCTT
URA3AICYC1Rnew	CCG-CCTGCAGG-TCATGTAATTAGTTATGTCACGCTTACA
URA3AIIDP1F	CCG-GAATTC-GATGGTAATGATCCGAACTTGG
URA3AIIDP1Rnew	CCG-CCTGCAGG-TCGAATTTACGTAGCCCAATCTAC
URA3AIPRM9F	CCG-GAATTC-ATTTTCAACATCGTATTTTCCGAAGC
URA3AIPRM9Rnew	CCG-CCTGCAGG-ACAGAAGACGGGAGACACT
GREcontrol-fwd	GTCTGTTATTAATTTCACAGGTAGTTCTGGTCCATTGGTGAAAGTTTGCGGCTTGCAGAGCACAGAGGCCGCAGAATGT-GTGATGACAAAACCTCTTCCG
TART-BB-f	T-CAAACAGGAATCGAATGCAACCG
TART-BB-r	AATTGTCCTTTTAACAGCGATCGC
ip-G418f	T-GAAAGAAATGCATAAGCTTTTGC
ip-G418r	AGACTTGTTCAACAGGCC
G418-stop-f	CGAATTCACTGGCCGTCG
G418-stop-r	CAAGCTTGGCGTAATCATGG
G418-stop2-F	GACGTTGTAAAACGACGGCCAGTGAATTCGTCTGTAGCCCTCAACGGAAA
G418-stop2-r	AGGTCCGCCGGCGTTGGACGAGCGGGTGCAACTAATTGACGGGAGT
JMW-0077	CGAATTCACTGGCCGTCG
JMW-0078	CAAGCTTGGCGTAATCATGG
JMW-0092	GACGTTGTAAAACGACGGCCAGTGAATTCGTCTGTAGCCCTCAACGGAAA

JMW-0093	AGGTCCGCCGGCGTTGGACGAGCGGGTGCAACTAATTGACGGGAGT
JMW-0094	TGCACCCGCTCGTCCAACGCCGGCGGACCTCGGTGCAACACTACTTCAACTTC
JMW-0095	ACAGCTATGACCATGATTACGCCAAGCTTGTTGTTCTCCTCATTCCAGGTG
JMW-0101	TCTGTAGCCCTCAACGGAAA
JMW-0102	TTGTTCTCCTCATTCCAGGTG
JMW-0163	TGACACCGATTATTTAAAGCTGCAG
JMW-0164	CTTTGCCTTCGTTTATCTTGCC
JMW-0165	TATATATATCGTATGCTGCAGCTTTAAATAATCGGTGTCAGCCTGATCTGTTTAGCTTGC
JMW-0166	TATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAGACTGGATGGCGGCGTTAGTA
JA-211	AAACTTAGATTAGATTGCTATGC
JA-240	TAAAGCTGCAGCCCAA
JMW-0210	ATTCATTAGATCGATAAGCTTGG
JMW-0211	TTGCTCATTAGAAAGAAAGCAT
JMW-0225	TAGCCAGGAGGAGAGTGACGGCTCCTATGTTGTGTGG
JMW-0226	ATGACGGAAGATGCTCGGGCGGGGTATCGTATGCTTC
JMW-0227	GCGGAACCCCTATTTGTTTATT
JMW-0228	GCTGTTTCCTGTGTGAAATTGTT
JMW-0229	TTGTGAGCGGATAACAATTTCACACAGGAAACAGCTTTGCTTCTCTCTTCTTCCTGTG
JMW-0230	CACACAACATAGGAGCCGTCACTCTCCTCCTGGCTATTG
JMW-0231	AAGCATACGATACCCCGCCCGAGCATCTTCCGTCATTA
JMW-0232	GTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCCTTTTAGGCCCAGAATTAGG
JMW-0280	TTTGCTTCTCTCTTCTTCCTGTG
JMW-0281	TCCACAACCAGTTTCATTCG

Table A-2: Primers used in Chapter 3 (IDT)

Primer Name	Sequence
Quikchange multi - Ty1 RT AA226 round two	gtatttaaaaacagtcaagtgacaatttgtttattc**vrw**gatgatatggtattgtttagcaaaaatctaaattca
Quikchange multi - Ty1 RT AA145 round two	cctgtcacttgcattagacaataactactatattacacaa**wdk**gacatatcttcggcatatttg
Quikchange multi - Ty1 RT AA225	gtatttaaaaacagtcaagtgacaatttgttta**ndg**gtagatgatatggtattgtttagcaaa

round two	aatct
Quikchange - Ty1 RT AA226 -> CAG	tatttaaaaacagtcaagtgacaatttgtttattc**CAG**gatgatatggtattgtttagcaaaaatctaaattc
ARTrev	TTATGCAATCAGGTGAATACGTTTCTT
Ty2600F	GTGCAAGT-AGTCGCTGAACGGCTAAAC
Ty2600R	GCGTGCA-CCATGTGCTCGGGAATCC
QMTy1RTL145Sf	cttgcattagacaataactactatattacacaatctgacatatcttcggcatatttgtat
QMTy1RTL145Sr	atacaaatatgccgaagatatgtcagattgtgtaatatagtagttattgtctaatgcaag
Ty1RTL151Af	GC-TTTGTATGCAGACATCAAAGA
Ty1RTL151Ar	TGCCGAAGATATGTCTAATTG
Ty1RTK93Rf	AAAAAG-GACGGTACTCATAAAGCTAG
Ty1RTR94Kf	A**G**ACGTGACGGTACTCATAAAGC
Ty1RTK93R94rev	CTTGTTGAAGATAAACATTGAGT
YFPfusR	ATTCCAACATACCACCCATAATGTAATAGATC-ttatttgtacaattcatccataccatgg
YFPHARTfusF	AAGTTGGTTTCTGCCGGTATTAGAAAGATCTTG-tctaaaggtgaagaattattcactggt
YFPTARTfusF	ACATTTAAACTATTAACTAACAAATGGATTCAT-tctaaaggtgaagaattattcactggt
YFPlinkerF	ggtgacggtgctggtttaattaac-tctaaaggtgaagaattattcactggt
Ty1RTlinkerR	ATGAATCCATTTGTTAGTTAATAGTTTAAATGTTT
HIVRTlinkerR	CAAGATCTTTCTAATACCGGCAGA
EcoRITkc1F	G-gaattc-TTTGAAAGATGATACTCTTTATTTCTAGACAG
SbfITkc1R	GG-cctgcagg-TATATATATATAACTGTCTAGAAATAAAGAGTATCATCTT
EcoRITkc5F	G-gaattc-TTTGAAAGATGATACTCTTTATTTATATATATATATATATATATATA
SbfITkc5R	GG-cctgcagg-TATATATATATATATATATATATAAATAAAGAGTATCATCTTTCAAA
EcoRITkc8SbfIF	G-gaattc-TTTGAAAAAATTTATTAAAAAAAAAATATATA-cctgcagg-CC
SbfITkc8EcoRIR	GG-cctgcagg-TATATATTTTTTTTTTAATAAATTTTTTCAAA-gaattc-C
EcoRITkc22SbfIF	G-gaattc-TTTTTAGATGATACTCTTTATTTCTAGACAGTTATATA-cctgcagg-CC
SbfITkc22EcoRIR	GG-cctgcagg-TATATAACTGTCTAGAAATAAAGAGTATCATCTAAAAA-gaattc-C
miniTy1-f	atcgataagcttgggCTG
miniTy1-r	GTTAACATTGGTGGTGGTC
HelperTy1-f	aattcgccctatagtga
HelperTy1-r	cCTAATGAATCCATTTGTTAG

Aintron-f	GTATGTTAATATGGACTAAAGG
Aintron-r	CAAATGGTATTATTTATAACAG
TART-term-f	cc-gaattc-tacccaattcgccctatagt
TART-term-r	gg-cctgcagg-ccTGAGAAATGGGTGAATG
SI-AscI-f	GTATGTTAATATGGACTAAAGGAGGCTTTTGGCGCGCCGAATTTTTACTAACAAATGGTATTATTTATAACAG
SI-AscI-r	CTGTTATAAATAATACCATTTGTTAGTAAAAATTCGGCGCGCCAAAAGCCTCCTTTAGTCCATATTAACATAC
URA3-ATG-SIf	CC-GGCGCGCCGAATTTTTACTAACAAATGGTATTATTTATAACAG-taaacttagattagattgctatgc
URA3-ATG-Sir	CC-GGCGCGCCAAAAGCCTCCTTTAGTCCATATTAACATAC-tgtcgaaagctacatataagg
URA3-ATG-If	CC-CCCGGGTACCGAGCTCGAATTTTTACTAACAAATGGTATTATTTATAACAG-taaacttagattagattgctatgc
URA3-ATG-Ir	CC-CCCGGGGATCCTCTAGAGTCGACCTGCAGAAAAGCCTCCTTTAGTCCATATTAACATAC-tgtcgaaagctacatataagg
URA3-ATG-SIf2	CC-GGCGCGCCGAATTTTTACTAACAAATGGTATTATTTATAACAG-tCTTTGCCTTCGTTTATCTTG
URA3-ATG-If2	CC-CCCGGGTACCGAGCTCGAATTTTTACTAACAAATGGTATTATTTATAACAG-tCTTTGCCTTCGTTTATCTTG
TART-term2-r	gg-gaattcggg-cctgcagg-ccTGAGAAATGGGTGAATG
DHH1f	CC-CCCGGG-ATGGGTTCCATCAATAAT
DHH1r	CC-CTCGAG-TTAATACTGGGGTTGTG
LSM1f	CC-CCCGGG-ATGTCTGCAAATAGCA
LSM1r	CC-CTCGAG-TTAGTACATGTCAGATTTATG
PAT1f	CC-CCCGGG-ATGTCCTTCTTTGGGTTAG
PAT1r	CC-CTCGAG-TTACTTTAGTTCTGATATTTCACC
SacI-HSXf	CC-gagctc-GGGTTCATTACAGCAGC
PstI-HSXr	CC-ctgcag-TATATTTACCATCAACTCCGC
Ty1INrev	TGATCATTGACGATAACAGGTTGAC
Ty1INfwd	TGTGAGAAATTCACTAGCTTCACC
TART-BB-f	aaacttagattagattgctatgc
TART-BB-r	CACCGATTATTTAAAGCTGC
ip-PRM9-f	AATGGATTCATTAGg-ACAGAAGACGGGAGACAC
ip-PRM9-r	actatagggcgaatt-ATTTTCAACATCGTATTTTCCGAAG
ip-CYC-f	AATGGATTCATTAGg-attagttatgtcacgcttaca
ip-CYC-r	actatagggcgaatt-caaattaaagccttcgagc

ip-G418f	TTTAAATAATCGGTG-ttagaaaaactcatcgagcatc
ip-G418r	ATCTAATCTAAGTTT-atgggtaaggaaaagactcac
G418-stop-f	T-Caaacaggaatcgaatgcaaccg
G418-stop-r	aattgtccttttaacagcgatcgc
Integrase-end-f	GGTATCACGAGACACATGG
Cyc-beg-f	attagttatgtcacgcttacattcac
GAL1f	cgcgcaattaaccctcactaaagggaacaaaagctggagc-AGTACGGATTAGAAGCCGCC
GAL1r	TATAGTTTTTTCTCCTTGACGTTAAAGTATAG
T7RNAP-f	ATATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATA-atgaacacgattaacatcgc
T7RNAP-r	gggagggcgtgaatgtaagcgtgacataactaat-ttacgcgaacgcgaag
5'LTRf	CTCTACCGCCTCGAGGAGAAC
beforeprom-rev	gagctccagcttttgttccc
T7p-f	gcgcgcaattaaccctcactaaagggaacaaaagctggagctc-gccgggaatttaatacg
T7p-r	AGGTATACAGAATATACTAGAAGTTCTCCTCGAGGCGGTAGAG-tctccctatagtgagtc

Table A-3: Primers used in Chapter 4 (IDT)

Primer Name	Sequence
T7p-Spinach2-Fwd	GAAATTAATACGACTCACTATA-GATGTAACTGAATGAAATGG
Spinach2r	GATGTAACTAGTTACGGA

Table A-4: Primers used in Chapter 5 (IDT)

Primer Name	Sequence
EcoRI-ARO4-wt-r	CCC-gaattc-CTATTTCTTGTTAACTTCTCTTCTTTG
XmaI-ARO4-f	CGC-cccggg-ATGAGTGAATCTCCAATGTTCG
ARO4-reg-f	GCACAGAGAATTGGCCTC
ARO4-reg-r	GCAATTGAGCCTTAGCTTCTG
ext-ARO4-reg-f	GGTGCCAGAACCACCGAATCTCAACT-GCACAGAGAATTGGCCTC
ext-ARO4-reg-r	ATCATTAGACCGTTGGAACCGGCAG-GCAATTGAGCCTTAGCTTCTG
ARO4-reg-BB-f	GCTAAGGCTCAATTGCCTGCC
ARO4-reg-BB-r	CCAATTCTCTGTGCAGTTGAGATTCG
H230Y-ins-f	GGTGCCAGAACCACCGAATC
H230Y-ins-r	ACCACCTCTTAGAATAACGAAGCAG

H230Y-BB-f	ACACTGCTTCGTTATTCTAAGAGGTGG
ARO4-reg-BB-r	CCAATTCTCTGTGCAGTTGAGATTCG

Table A-5: Primers used in Chapter 6 (IDT)

Primer Name	Sequence
P2A-f	TCCTGGATTTTCTTCAACATCACCATCTTGTTTTAATAAAGAAAAGTTGGTAGCACCAGA
P2A-r	TCTGGTGCTACCAACTTTTCTTTATTAAAACAAGATGGTGATGTTGAAGAAAATCCAGGA
P2Av3-f	TCCTGGGTTCTCCTCGACGTCTCCGTCCTGCTTCAACAATGAGAAATTAGTTGCTCCTGA
P2Av3-r	TCAGGAGCAACTAATTTCTCATTGTTGAAGCAGGACGGAGACGTCGAGGAGAACCCAGGA
XmaI-YFPf	CC-CCCGGG-ATG-tctaaaggtgaagaattattca
EcoRI-YFPr	CC-gaattc-ttatttgtacaattcatccatac
SecG-YFPf	ACCAGTGATTTCAACAACAGTGGGTTCAGCTGCTGAAGGTTCATTGGATAAGAGAGAAGCT-tctaaaggtgaagaattattcac
SecG-YFPr	TGAGCCAAAGCCAATGGCAAAGCAGCGAAAATAGCCAACAACACAATCAACACCTT-CATcccgggggatc
SecG-P2A-f	AGATGGTGATGTTGAAGAAAATCCAGGACCA-tctaaaggtgaagaattattcac
SecGP2Ad-f	AGATGGTGATGTTGAAGAAAATCCAGGATTT-tctaaaggtgaagaattattcac
SecGP2A-r	TGTTTTAATAAAGAAAAGTTGGTAGCACCAGAACC-AGCTTCTCTCTTATCCAATG
E2Af	AATTAGCCGGTGATGTTGAAAGTAACCCAGGACCA-tctaaaggtgaagaattattcac
E2Ar	TTAATAAAGCATAGTTGGTACATTGACCAGAACC-AGCTTCTCTCTTATCCAATG
F2Ar	TCAACAAATCGAAATTTAAAGTTTGCTTAACACCAGAACC-AGCTTCTCTCTTATCCAATG
T2Af	TTGTGGTGATGTTGAAGAAAATCCAGGACCA-tctaaaggtgaagaattattcac
T2Ar	GTTAATAAAGAGCCACGACCTTCACCAGAACC-AGCTTCTCTCTTATCCAATG
P2Ad-f	TTT-tctaaaggtgaagaatta
P2Ad-r	TCCTGGATTTTCTTCAACA
SecG-E2Af	GCTGAAGGTTCATTGGATAAGAGAGAAGCT-GGTTCTGGTCAATGTACC

SecG-E2Ar	ggacaacaccagtgaataattcttcacctttaga-TGGTCCTGGGTTACTTTC
SecG-YFP-BBf	tctaaaggtgaagaattattc
SecG-YFP-BBr	AGCTTCTCTCTTATCCA
2ArandBB-f	GGTGATGTTGAAGAAAATCC
2ArandBB-r	ACCAGAACCAGCTTCTC
rand2A	GTTCATTGGATAAGAGAGAAGCTGGTTCTGGT-NNNNNNNNNNNNNNNNNNNNNNNNN-GGTGATGTTGAAGAAAATCCAGGACCAtcta
rand2A-r	tagaTGGTCCTGGATTTTC
rand2A-f	GTTCATTGGATAAGAGAGAAGC
rand2A-ext-f	AGTGATTTCAACAACAGTGGGTTCAGCTGCTGAAG-GTTCATTGGATAAGAGAGAAGC
rand2A-ext-r	caaaattgggacaacaccagtgaataattcttcacctt-tagaTGGTCCTGGATTTTCTTC
T2A-beg-f	GGTTCTGGTGAAGGTCG
2A-end-rev	TGGTCCTGGATTTTCTTCAAC
T2A-mid-r	ACAAGTTAATAAAGAGCCACG
2ArandBB2-F	GGTGATGTTGAAGAAAATCCAGGAC
2ArandBB2-r	GATAAGAGAGAAGCTGGTTCTGGT
r2Ad-s-f	GCTGAAGGTTCATTGGATAAGAGAGAAG
r2Ad-s-r	aacaccagtgaataattcttcacctttaga-AGCTTCTCTCTGTCCTGGATTTTC
P2A-f	GGTTCTGGTGCTACC
d-P2A-r	GTTTTAATAAAGAAAAGTTGGTAGCACCAGAACC-AGCTTCTCTCTGTCCTGGATTTTC
E2A-f	GGTTCTGGTCAATGTAC
d-E2A-r	TTTTAATAAAGCATAGTTGGTACATTGACCAGAACC-AGCTTCTCTCTGTCCTGGATTTTC
F2A-f	GGTTCTGGTGTTAAGC
d-F2A-r	CAAATCGAAATTTAAAGTTTGCTTAACACCAGAACC-AGCTTCTCTCTGTCCTGGATTTTC
2Arandlib2-F	GTTCTGGT-NNNNNNNNNNNNNNNNNNNNNNNNN-GGTGATGTTGAAGAAAATCCAGGAC
2Arandlib2-r	AGGTTCATTGGATAAGAGAGAAGCTG
XylA-f	tcgacggattctagaactagtggatcccccgggctgcagg-ATGGCTAAAGAATATTTCCC
XylA-r	TTGATACATCGCGACAATAG
XP-P2A-f	GGAAAACAAGAACTATACGAGGCTATTGTCGCGATGTATCAA-GGTTCTGGTGCTACCAAC
XP-P2A-r	GGAAACCTCTCTTGTCTGTCTCTGAATTACTGAACACAA-TGGTCCTGGATTTTCTTCAAC

XP-P2Ad-r	GGAAACCTCTCTTGTCTGTCTCTGAATTACTGAACACAA-AAATCCTGGATTTTCTTCAAC
XKS1-f	TTGTGTTCAGTAATTCAGAG
XKS1-r	tctcgaggtcgacggtatcgataagcttgatatcgaatt-TTAGATGAGAGTCTTTTCCAG
2Arandlib2-r-fix	CAGCTTCTCTCTTATCCAATGAACC
IF-2ABB-f	gttgtcccaattttggttg
IF-2ABB-r	GTTGTTGAAATCACTGGTTG
BC-bb-f	gattacgcgaacatcctc
BC-bb-r	ctgcccttcccatcc
CrtI-f	ggaaaagaacaagatcagg
CrtI-r	gaaagcaagaacaccaac
CrtYB-d12-F	gaaagtcttgagtgtggtcatgagcggatgggaagggcag-GGTTCTGGTATGAACATAGG
CrtYB-d8-f	cggaaagtcttgagtgtggtcatgagcggatgggaagggcag-GGTTCTGGTGTTCACATG
CrtI-1-r	cacgatgatagctgtgggtttatcctgatcttgttcttttcc-TGGTCCTGGATTTTCTTC
CrtI-2-r	cacgatgatagctgtgggtttatcctgatcttgttcttttcc-TGGTCCTGGGTTACTTTC
CrtI-d12-F	tggggccgtgatcgctcgatccgttggtgttcttgctttc-GGTTCTGGTATGAACATAGG
CrtI-d8-f	attggggccgtgatcgctcgatccgttggtgttcttgctttc-GGTTCTGGTGTTCACATG
CrtE-1-r	agtaaactcgagtggaattgctgtgaggatgttcgcgtaatc-TGGTCCTGGATTTTCTTC
CrtE-2-r	agtaaactcgagtggaattgctgtgaggatgttcgcgtaatc-TGGTCCTGGGTTACTTTC
CrtYB-T2A-f	cggaaagtcttgagtgtggtcatgagcggatgggaagggcag-AGAGCAGAAGGAAGGG
CrtI-T2A-r	cacgatgatagctgtgggtttatcctgatcttgttcttttcc-TGGTCCTGGATTCTCC
CrtI-T2A-f	gattggggccgtgatcgctcgatccgttggtgttcttgctttc-AGAGCTGAGGGTAGAGG
CrtE-T2A-r	agtaaactcgagtggaattgctgtgaggatgttcgcgtaatc-AGGACCAGGGTTTTCTTC
T2A1-f	AGAGCAGAAGGAAGGGGTTCTTTGTTGACTTGTGGAGATGTTGAGGAGAATCCAGGACCA
T2A1-r	TGGTCCTGGATTCTCCTCAACATCTCCACAAGTCAACAAAGAACCCCTTCCTTCTGCTCT
T2A2-F	AGAGCTGAGGGTAGAGGAAGTCTACTAACATGCGGTGACGTAGAAGAAAACCCTGGTCCT
T2A2-r	AGGACCAGGGTTTTCTTCTACGTCACCGCATGTTAGTAGACTTCCTCTACCCTCAGCTCT

BCseq1	tgaccgacttcctaccgagg
BCseq2	tcactctccgtaccatatggaa
SecG-T2A1f	AGTGGGTTCAGCTGCTGAAGGTTCATTGGATAAGAGAGAAGCT-AGAGCAGAAGGAAGGGG
YFP-T2A1r	caaaattgggacaacaccagtgaataattcttcacctttaga-TGGTCCTGGATTCTCCTC
SecG-T2A2f	AGTGGGTTCAGCTGCTGAAGGTTCATTGGATAAGAGAGAAGCT-AGAGCTGAGGGTAGAGG
YFP-T2A2r	caaaattgggacaacaccagtgaataattcttcacctttaga-AGGACCAGGGTTTTCTTC
BBf	aattcgatatcaagcttatcg
BBr	TTGATACATCGCGACAATAG
Insert (transport)f	TTGTTGACTTGTGGAGATGTTGAGGAGAATCCAGGACCA-ggtttggaggacaatagaatg
Insert (transport)r	aatctcgaggtcgacggtatcgataagcttgatatcgaattcta-agatggaacgaagccc
Insert (T2A1)f	AGGAAAACAAGAACTATACGAGGCTATTGTCGCGATGTATCAA-AGAGCAGAAGGAAGGGG
Insert (T2A1)r	aacgttgacgaaacgcttaaccattctattgtcctccaaacc-TGGTCCTGGATTCTCCTC
BB2r	AAATCCTGGATTTTCTTCAAC
Insertf	TAAAACAAGATGGTGATGTTGAAGAAAATCCAGGATTT-ggtttggaggacaatagaatgg
Insert(trans)r	TCAACATCTCCACAAGTCAACAAAGAACCCCTTCCTTCTGCTCT-agatggaacgaagccc
Insert (T2A1 beta)f	gcatgtcagcaaggcttggaagtctaagggcttcgttccatct-AGAGCAGAAGGAAGGGG
Insert (T2A1 beta)r	GTTGGAAACCTCTCTTGTCTGTCTCTGAATTACTGAACACAA-TGGTCCTGGATTCTCCTC
Insert (Xks)f	TTTGTTGACTTGTGGAGATGTTGAGGAGAATCCAGGACCA-TTGTGTTCAGTAATTCAGAG
Insert (Xks)r	tctcgaggtcgacggtatcgataagcttgatatcgaatt-TTAGATGAGAGTCTTTTCCAG
Insert(XKs1)f	gagcatgtcagcaaggcttggaagtctaagggcttcgttccatct-GGTTCTGGTGCTACCAAC
Insert(XKs1)r	aatctcgaggtcgacggtatcgataagcttgatatcgaatt-TTAGATGAGAGTCTTTTCC
Insert (trans)r	TCACCATCTTGTTTTAATAAAGAAAAGTTGGTAGCACCAGAACC-agatggaacgaagccc
Insert (P2Ad beta)f	gagcatgtcagcaaggcttggaagtctaagggcttcgttccatct-GGTTCTGGTGCTACC
Insert (P2Ad beta)r	GGAAACCTCTCTTGTCTGTCTCTGAATTACTGAACACAA-AAATCCTGGATTTTCTTCAAC

Trans-r-fix	AGATGGAACGAAGCCC
BB-XKS1-f	TTGTGTTCAGTAATTCAGAGACAG
CiGxs-f	ggtttggaggacaatagaatg
Insert(XKs1)r-fix	cgaggtcgacggtatcgataagcttgatatcgaatt-TTAGATGAGAGTCTTTTCCAGTTC
P2Ad2-trans-f	AACAAGAACTATACGAGGCTATTGTCGCGATGTATCAA-GGATCAGGAGCAACTAATTTC
P2Ad2-trans-r	ttgacgaaacgcttaaccattctattgtcctccaaacc-AAAACCAGGGTTCTCCTC
P2Ad-trans-f	tcagcaaggcttggaagtctaagggcttcgttccatct-GGTTCTGGTGCTACCAAC
P2Ad-trans-r	AACCTCTCTTGTCTGTCTCTGAATTACTGAACACAA-AAATCCTGGATTTTCTTCAACATC
T2A-trans-f	AACAAGAACTATACGAGGCTATTGTCGCGATGTATCAA-GGTTCTGGTGAAGGTCGTG
T2A-trans-r	ttgacgaaacgcttaaccattctattgtcctccaaacc-TGGTCCTGGATTTTCTTCAAC
P2A-trans-r	AACCTCTCTTGTCTGTCTCTGAATTACTGAACACAA-TGGTCCTGGATTTTCTTCAAC
2A-ep-r	caccagtgaataattcttcacctttaga
PAD1-a-f	aagttttctagaactagtggatcccccgggctgcagg-ATGCTCCTATTTCCAAGAAGAAC
PAD1-a-r	CTTGCTTTTTATTCCTTCCCAAC
PAD1-b-f	CTCCTATTTCCAAGAAGAACTAATATAGCC
PAD1-b-r	tcgaggtcgacggtatcgataagcttgatatcg-TTACTTGCTTTTTATTCCTTCCCAACG
FDC1-a-f	AGGAAGCTAAATCCAGCTTTAG
FDC1-a-r	aggtcgacggtatcgataagcttgatatcg-TTATTTATATCCGTACCTTTTCCAATTTTC
FDC1-b-f	gttttctagaactagtggatcccccgggctgcagg-ATGAGGAAGCTAAATCCAGCTTTAG
FDC1-b-r	TTTATATCCGTACCTTTTCCAATTTTCATTTAC
P2A-PF-a-f	CACGCTGACACTTTTCCTCGTTGGGAAGGAATAAAAAGCAAG-GGTTCTGGTGCTACCAAC
P2A-PF-a-r	GATAAAGTCTCTAAATTCTAAAGCTGGATTTAGCTTCCT-TGGTCCTGGATTTTCTTCAAC
P2Ad-PF-a-r	GATAAAGTCTCTAAATTCTAAAGCTGGATTTAGCTTCCT-AAATCCTGGATTTTCTTCAAC
P2A-PF-b-f	GTTGACAAAGTAAATGAAAATTGGAAAAGGTACGGATATAAA-GGTTCTGGTGCTACCAAC
P2A-PF-b-r	TTTGAAAAAGGCTATATTAGTTCTTCTTGGAAATAGGAG-TGGTCCTGGATTTTCTTCAAC
P2Ad-PF-b-r	TTTGAAAAAGGCTATATTAGTTCTTCTTGGAAATAGGAG-

	AAATCCTGGATTTTCTTCAAC
d8/12-2A-f	CCAtctaaaggtgaagaattattc
d8/12-2A-r	TCCTGGATTTTCTTCAACATC
KanMX-qPCR-f	ggtgaaaatattgttgatgcgctggc
KanMX-qPCR-r	attcgtgattgcgcctgagc
ARO4-qPCR-f	CCCTTAGTAGTGGTGATAGCAGCA
ARO4-qPCR-r	GGTACCATCGGTACCGTTCTT
SecG-BB-r	AGCTTCTCTCTTATCCAATGAACC
F6-BB-f	GGTTCTGGTCTTCAGCTGATTC
F9-BB-f	GGTTCTGGTTTTTGTACATCAGC
J3-BB-f	GGTTCTGGTTTTCTTGGATGG
J6-BB-f	GGTTCTGGTCTATCTTGGCTC
d12-I6-r	CATCACCCCCTCGGTAGGCTAGAATCAGCTGAAGACCAGAACC-AGCTTCTCTCTGTCCTG
d12-I9-r	CATCACCCCAGTAACGTCCAGCTGATGTACAAAAACCAGAACC-AGCTTCTCTCTGTCCTG
d12-G3-r	CATCACCCTTAAAATGGTGTTTCCATCCAAGAAAACCAGAACC-AGCTTCTCTCTGTCCTG
d12-G6-r	CATCACCGTATATCCCCTGGAAGAGCCAAGATAGACCAGAACC-AGCTTCTCTCTGTCCTG
Xyl2r-fix	ctcgaggtcgacggtatcgataagcttgatatc-ttactcagggccgtcaatg
XylA-f	CGACGGATA-TGGCTAAAGAATATTTCCCTC
XylA-r	GTTGGTAGCACCAGAACCTTGATACATCGC
XKS1-r	TTAGATGAGAGTCTTTTCCAGTTCGCTTAAGG
d12-PF-f	CGCTGACACTTTTCCTCGTTGGGAAGGAATAAAAAGCAAG-GGTTCTGGTATGAACATAGG
d8-PF-f	CACGCTGACACTTTTCCTCGTTGGGAAGGAATAAAAAGCAAG-GGTTCTGGTGTTCACATG
P,T2A-PF-r	CTGGATAAAGTCTCTAAATTCTAAAGCTGGATTTAGCTTCCT-TGGTCCTGGATTTTCTTC
T2A-PF-f	CACGCTGACACTTTTCCTCGTTGGGAAGGAATAAAAAGCAAG-GGTTCTGGTGAAGGTCG
F2A-PF-f	CGCTGACACTTTTCCTCGTTGGGAAGGAATAAAAAGCAAG-GGTTCTGGTGTTAAGCAAAC
E2A-PF-f	CACGCTGACACTTTTCCTCGTTGGGAAGGAATAAAAAGCAAG-GGTTCTGGTCAATGTACC
F,E2A-PF-r	CTGGATAAAGTCTCTAAATTCTAAAGCTGGATTTAGCTTCCT-TGGTCCTGGGTTACTTTC
XmaI-ARO3t-f	ACA-CCCGGG-ATG-CCAATGTTCATTAAAAACGATCAC
XhoI-ARO3t-r	CGC-ctcgag-CTA-TCCTGGATTTTCTTCAACATC

XhoI-ARO4-r	CCC-CTCGAG-CTATTTCTTGTTAACTTCTCTTCTTTGTC
dp-TEFf	gcgcaattaaccctcactaaagggaacaaaagctgg-atagcttcaaaatgtttctactcc
dp-CYCr	tggaagagtaaaaaaggagtagaaacattttgaagctatg-caaattaaagccttcgagcg
SacI-TEFf-dp	CCC-gagctc-atagcttcaaaatgtttctactcc
SacI-CYCr-dp	CCC-gagctc-caaattaaagccttcgagcg
d12-PF-f	AAAACAAGAACTATACGAGGCTATTGTCGCGATGTATCAA-GGTTCTGGTATGAACATAGG
d8-PF-f	GGAAAACAAGAACTATACGAGGCTATTGTCGCGATGTATCAA-GGTTCTGGTGTTCACATG
P,T2A-PF-r	aacgttgacgaaacgcttaaccattctattgtcctccaaacc-TGGTCCTGGATTTTCTTC
T2A-PF-f	GGAAAACAAGAACTATACGAGGCTATTGTCGCGATGTATCAA-GGTTCTGGTGAAGGTCG
XmaI-XylA-f	ACA-CCCGGG-ATGGCTAAAGAATATTTCCCTC
XhoI-XylA-r	CCC-CTCGAG-CTATTGATACATCGCGACAATAG
SacI-GPDf-dp	CCC-gagctcagtttatcattatcaatactc
XmaI-CiGXS-f	ACA-CCCGGG-ATGggtttggaggacaatag
XhoI-CiGXS-r	CCC-CTCGAG-CTAagatggaacgaagccc
ARO2f	CCATCTTGTTTTAATAAAGAAAAGTTGGTAGCACCAGAACC-ATGAACCACGGATCTGGA
ARO2r	ttgctcattagaaagaaagcatagcaatctaatctaagttt-ATGTCAACGTTTGGGAAAC
ARO3f	GTCCTGCTTCAACAATGAGAAATTAGTTGCTCCTGATCC-TTTTTTCAAGGCCTTTCTTC
ARO3r	AAAACAAGATGGTGATGTTGAAGAAAATCCAGGACCA-ATGTTCATTAAAAACGATCACG
ARO4f	atcgataagcttgggCTGCAGCTTTAAATAATCGGTG-CTATTTCTTGTTAACTTCTCTTC
ARO4r	TTGAAGCAGGACGGAGACGTCGAGGAGAACCCAGGACCA-ATGAGTGAATCTCCAATGTTC
ARO2KOf	ATAGTATCGAAAAAAGAAAATAGTATCATAGCACAGAGGC-tgcaggtcgacaacccttaa
ARO2KOr	TAAATATACATAACTCTTGAGGGGTTTTGTTTCTATCT-ggccactagtggatctgatatc

Appendix B – Sequences

Table B-1: Sequences of Aptamer Templates in Chapter 4

Aptamer Name	Sequence
Tyr1	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAGGCAGT CAACTCGTGCGATCGTGAAAACGGGGCAAGATGGCCTTACAG CGGTCAATACGGGGGTCATCAGATAGGGAGGCCTCCTGGTTT GTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1b	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAgcttgtagg gGGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCCTC CTGGTTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1c	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAgtacaggg GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCCTC CTGGTTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1d	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAtttgtacagg gGGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCCTC CTGGTTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1e	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAgtaGaggg GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCCTC CTGGTTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1f	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAgtacTggg GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCCTC CTGGTTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1g	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAgtaGTggg GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCCTC CTGGTTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1h	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAgtacaggg GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCCTC CTcGTTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1i	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAgtacaggg GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCCTC CAGGTTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1j	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAgtacaggg GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCCTC CACGTTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1M-1	GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCC
Tyr1M-2	GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCC

Tyr1M-3	GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCC
Tyr1M-4	GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCC
Tyr1M-5	GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCC
Tyr1M-6	GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCC
Tyr1M-7	GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCC
DNT1	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCATTCCAG CTCGGTACCATAACACAAGTGGTAGACTATTCTCTGGTACGTG CGCCCCCGGCCGTATTACGGGAGCACGCCGGCTAAGGGATG TTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC

Table B-2: Sequences of Aptamer Templates in Chapter 5

Aptamer Name	Sequence
Thrombin	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCAGGAACA AAGCTGAAGTACTTACCTTGTTGAGTAGAGTGTGAGCTCCGTA ACTAGTTACATC
Streptavidin	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCACGACCG ACCAGAATCATGCAAGTGCGTAAGATAGTCGCGGGCCGGGGT TGTTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
DNT1	GGATGTAACTGAATGAAATGGTGAAGGACGGGTCCATTCCAG CTCGGTACCATAACACAAGTGGTAGACTATTCTCTGGTACGTG CGCCCCCGGCCGTATTACGGGAGCACGCCGGCTAAGGGATG TTGAGTAGAGTGTGAGCTCCGTAACTAGTTACATC
Tyr1M-1	GGCAGTCAACTCGTGCGATCGTGAAAACGGGGCAAGATGGC CTTACAGCGGTCAATACGGGGGTCATCAGATAGGGAGGCC
SAM	GACGCGACTGAATGAAATGGTGAAGGACGGGTCCACGAAAG GATGGCGGAAACGCCAGATGCCTTGTAACCGAAAGGGTTGTT GAGTAGAGTGTGAGCTCCGTAACTAGTCGCGTC

References

1. Jang, Y.-S., et al., *Enhanced Butanol Production Obtained by Reinforcing the Direct Butanol-Forming Route in Clostridium acetobutylicum.* mBio, 2012. **3**(5).
2. Steen, E., et al., *Metabolic engineering of Saccharomyces cerevisiae for the production of n-butanol.* Microbial Cell Factories, 2008. **7**(1): p. 36.
3. Hong, S.H. and S.Y. Lee, *Metabolic flux analysis for succinic acid production by recombinant Escherichia coli with amplified malic enzyme activity.* Biotechnology and Bioengineering, 2001. **74**(2): p. 89-95.
4. Curran, K.A., et al., *Metabolic engineering of muconic acid production in Saccharomyces cerevisiae.* Metabolic Engineering, 2013. **15**(0): p. 55-66.
5. Schirmer, A., et al., *Microbial Biosynthesis of Alkanes.* Science, 2010. **329**(5991): p. 559-562.
6. Curran, K.A. and H.S. Alper, *Expanding the chemical palate of cells by combining systems biology and metabolic engineering.* Metabolic Engineering, 2012. **14**(4): p. 289-297.
7. Pourmir, A. and T. Johannes, *Directed evolution: selection of the host organism.* Computational and Structural Biotechnology Journal, 2012. **2**(3).
8. Liu, L., H. Redden, and H.S. Alper, *Frontiers of yeast metabolic engineering: diversifying beyond ethanol and Saccharomyces.* Current Opinion in Biotechnology, 2013(0).
9. Lanza, A.M., N.C. Crook, and H.S. Alper, *Innovation at the intersection of synthetic and systems biology.* Current Opinion in Biotechnology, 2012. **23**(5): p. 712-717.
10. Blazeck, J. and H. Alper, *Systems metabolic engineering: Genome-scale models and beyond.* Biotechnology Journal, 2010. **5**(7): p. 647-659.
11. Lee, J.W., et al., *Systems metabolic engineering of microorganisms for natural and non-natural chemicals.* Nat Chem Biol, 2012. **8**(6): p. 536-546.
12. Arnold, F.H., *Design by directed evolution.* Accounts of Chemical Research 1998. **31**: p. 125-131.
13. Voigt, C.A., et al., *Protein building blocks preserved by recombination.* Nat Struct Mol Biol, 2002. **9**(7): p. 553-558.
14. Chen, F., et al., *Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection.* Proceedings of the National Academy of Sciences, 2010.
15. Yamashiro, K., et al., *Improvement of Bacillus circulans β-amylase activity attained using the ancestral mutation method.* Protein Engineering Design and Selection, 2010. **23**(7): p. 519-528.
16. Farinas, Edgardo T., et al., *Directed Evolution of a Cytochrome P450 Monooxygenase for Alkane Oxidation.* Advanced Synthesis & Catalysis, 2001. **343**(6-7): p. 601-606.
17. Lee, S.-M., T. Jellison, and H.S. Alper, *Directed Evolution of Xylose Isomerase for Improved Xylose Catabolism and Fermentation in the Yeast Saccharomyces*

cerevisiae. Applied and Environmental Microbiology, 2012. **78**(16): p. 5708-5716.
18. Young, E.M., et al., *A molecular transporter engineering approach to improving xylose catabolism in Saccharomyces cerevisiae.* Metab Eng, 2012. **14**.
19. Romero, P.A. and F.H. Arnold, *Exploring protein fitness landscapes by directed evolution.* Nat Rev Mol Cell Biol, 2009. **10**(12): p. 866-876.
20. Hibbert, E.G., et al., *Directed evolution of biocatalytic processes.* Biomolecular Engineering, 2005. **22**(1–3): p. 11-19.
21. Wang, C.-w., M.-K. Oh, and J.C. Liao, *Directed Evolution of Metabolically Engineered Escherichiacoli for Carotenoid Production.* Biotechnology Progress, 2000. **16**(6): p. 922-926.
22. Sun, L., et al., *Modification of Galactose Oxidase to Introduce Glucose 6-Oxidase Activity.* ChemBioChem, 2002. **3**(8): p. 781-783.
23. Koryakina, I., et al., *Poly Specific trans-Acyltransferase Machinery Revealed via Engineered Acyl-CoA Synthetases.* ACS Chemical Biology, 2012. **8**(1): p. 200-208.
24. Bastian, S., et al., *Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli.* Metabolic Engineering, 2011. **13**(3): p. 345-352.
25. Joo, H., Z. Lin, and F.H. Arnold, *Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation.* Nature, 1999. **399**(6737): p. 670-673.
26. Sun, L., et al., *Expression and stabilization of galactose oxidase in Escherichia coli by directed evolution.* Protein Engineering, 2001. **14**(9): p. 699-704.
27. Alper, H., et al., *Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production.* Science, 2006. **314**(5805): p. 1565-1568.
28. Tang, S.-Y., H. Fazelinia, and P.C. Cirino, *AraC Regulatory Protein Mutants with Altered Effector Specificity.* Journal of the American Chemical Society, 2008. **130**(15): p. 5267-5271.
29. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries.* Nat Biotech, 1997. **15**(6): p. 553-557.
30. Crameri, A., et al., *Molecular evolution of an arsenate detoxification pathway by DNA shuffling.* Nat Biotechnol, 1997. **15**.
31. Patnaik, R., et al., *Genome shuffling of Lactobacillus for improved acid tolerance.* Nat Biotech, 2002. **20**(7): p. 707-712.
32. Otte, B., et al., *Genome Shuffling in Clostridium diolis DSM 15410 for Improved 1,3-Propanediol Production.* Applied and Environmental Microbiology, 2009. **75**(24): p. 7610-7616.
33. Winkler, J. and K.C. Kao, *Harnessing recombination to speed adaptive evolution in Escherichia coli.* Metabolic Engineering, 2012. **14**(5): p. 487-495.
34. Portnoy, V.A., D. Bezdan, and K. Zengler, *Adaptive laboratory evolution - harnessing the power of biology for metabolic engineering.* Current Opinion in Biotechnology, 2011. **22**(4): p. 590-594.

35. Fox, R.J. and M.D. Clay, *Catalytic effectiveness, a measure of enzyme proficiency for industrial applications.* Trends in Biotechnology, 2009. **27**(3): p. 137-140.
36. Coelho, P.S., et al., *Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes.* Science, 2013. **339**(6117): p. 307-310.
37. Ben-David, M., et al., *Catalytic Metal Ion Rearrangements Underline Promiscuity and Evolvability of a Metalloenzyme.* Journal of Molecular Biology, 2013. **425**(6): p. 1028-1038.
38. Suzuki, Y., et al., *Enhancement of the latent 3-isopropylmalate dehydrogenase activity of promiscuous homoisocitrate dehydrogenase by directed evolution.* Biochemical Journal, 2010. **431**(3): p. 401-410.
39. Umeno, D. and F.H. Arnold, *Evolution of a Pathway to Novel Long-Chain Carotenoids.* Journal of Bacteriology, 2004. **186**(5): p. 1531-1536.
40. Tobias, A.V. and F.H. Arnold, *Biosynthesis of novel carotenoid families based on unnatural carbon backbones: A model for diversification of natural product pathways.* Biochim Biophys Acta., 2006. **1761**(2): p. 235-246.
41. Yoshikuni, Y., T.E. Ferrin, and J.D. Keasling, *Designed divergent evolution of enzyme function.* Nature, 2006. **440**(7087): p. 1078-1082.
42. Ji, X.-J., et al., *Cofactor engineering through heterologous expression of an NADH oxidase and its impact on metabolic flux redistribution in Klebsiella pneumoniae.* Biotechnology for Biofuels, 2013. **6**(1): p. 7.
43. Hasegawa, S., et al., *Engineering of Corynebacterium glutamicum for High Yield l-Valine Production under Oxygen Deprivation Conditions.* Applied and Environmental Microbiology, 2012.
44. Alberstein, M., M. Eisenstein, and H. Abeliovich, *Removing allosteric feedback inhibition of tomato 4-coumarate:CoA ligase by directed evolution.* The Plant Journal, 2012. **69**(1): p. 57-69.
45. Báez-Viveros, J.L., et al., *Metabolic engineering and protein directed evolution increase the yield of L-phenylalanine synthesized from glucose in Escherichia coli.* Biotechnology and Bioengineering, 2004. **87**(4): p. 516-524.
46. Atsumi, S. and J.C. Liao, *Directed Evolution of Methanococcus jannaschii Citramalate Synthase for Biosynthesis of 1-Propanol and 1-Butanol by Escherichia coli.* Applied and Environmental Microbiology, 2008. **74**(24): p. 7802-7808.
47. Alper, H., K. Miyaoku, and G. Stephanopoulos, *Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets.* Nat Biotechnol, 2005. **23**.
48. Blazeck, J. and H.S. Alper, *Promoter engineering: Recent advances in controlling transcription at the most fundamental level.* Biotechnology Journal, 2013. **8**(1): p. 46-58.
49. Nevoigt, E., et al., *Engineering of Promoter Replacement Cassettes for Fine-Tuning of Gene Expression in Saccharomyces cerevisiae.* Applied and Environmental Microbiology, 2006. **72**(8): p. 5266-5273.

50. Nevoigt, E., et al., *Engineering promoter regulation.* Biotechnology and Bioengineering, 2007. **96**(3): p. 550-558.
51. Kagiya, G., et al., *Generation of a strong promoter for Escherichia coli from eukaryotic genome DNA.* Journal of Biotechnology, 2005. **115**(3): p. 239-248.
52. Vee Aune, T.E., et al., *Directed evolution of the transcription factor XylS for development of improved expression systems.* Microbial Biotechnology, 2010. **3**(1): p. 38-47.
53. Lee, S.K., et al., *Directed Evolution of AraC for Improved Compatibility of Arabinose- and Lactose-Inducible Promoters.* Applied and Environmental Microbiology, 2007. **73**(18): p. 5711-5715.
54. Koyanagi, T., et al., *Hyperproduction of 3,4-Dihydroxyphenyl-L-alanine (L-Dopa) Using Erwinia herbicola Cells Carrying a Mutant Transcriptional Regulator TyrR.* Bioscience, Biotechnology, and Biochemistry, 2009. **73**(5): p. 1221-1223.
55. Alper, H. and G. Stephanopoulos, *Global transcription machinery engineering: A new approach for improving cellular phenotype.* Metabolic Engineering, 2007. **9**(3): p. 258-267.
56. Klein-Marcuschamer, D., et al., *Mutagenesis of the Bacterial RNA Polymerase Alpha Subunit for Improvement of Complex Phenotypes.* Applied and Environmental Microbiology, 2009. **75**(9): p. 2705-2711.
57. Zhang, H., et al., *Random mutagenesis of global transcription factor cAMP receptor protein for improved osmotolerance.* Biotechnology and Bioengineering, 2012. **109**(5): p. 1165-1172.
58. Klein-Marcuschamer, D. and G. Stephanopoulos, *Assessing the potential of mutational strategies to elicit new phenotypes in industrial strains.* Proceedings of the National Academy of Sciences, 2008. **105**(7): p. 2319-2324.
59. Xiong, A.-S., et al., *Expression and Function of a Modified AP2/ERF Transcription Factor from Brassica napus Enhances Cold Tolerance in Transgenic Arabidopsis.* Molecular Biotechnology, 2013. **53**(2): p. 198-206.
60. Farmer, W.R. and J.C. Liao, *Improving lycopene production in Escherichia coli by engineering metabolic control.* Nat Biotech, 2000. **18**(5): p. 533-537.
61. Zhang, F. and J. Keasling, *Biosensors and their applications in microbial metabolic engineering.* Trends in Microbiology, 2011. **19**(7): p. 323-329.
62. Reed, B., J. Blazeck, and H. Alper, *Evolution of an alkane-inducible biosensor for increased responsiveness to short-chain alkanes.* Journal of Biotechnology, 2012. **158**(3): p. 75-79.
63. Collins, C.H., J.R. Leadbetter, and F.H. Arnold, *Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR.* Nat Biotech, 2006. **24**(6): p. 708-712.
64. Matsushika, A., et al., *Ethanol production from xylose in engineered Saccharomyces cerevisiae strains: current state and perspectives.* Applied Microbiology and Biotechnology, 2009. **84**(1): p. 37-53.

65. Young, E.M., et al., *A molecular transporter engineering approach to improving xylose catabolism in Saccharomyces cerevisiae.* Metabolic Engineering, 2012. **14**(4): p. 401-411.
66. Badran, A.H. and D.R. Liu, *In vivo continuous directed evolution.* Current Opinion in Chemical Biology, 2015. **24**: p. 1-10.
67. Esvelt, K.M., J.C. Carlson, and D.R. Liu, *A system for the continuous directed evolution of biomolecules.* Nature, 2011. **472**(7344): p. 499-503.
68. Dickinson, B.C., et al., *Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution.* Proceedings of the National Academy of Sciences, 2013. **110**(22): p. 9007-9012.
69. Carlson, J.C., et al., *Negative selection and stringency modulation in phage-assisted continuous evolution.* Nat Chem Biol, 2014. **10**(3): p. 216-222.
70. Badran, A.H., et al., *Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance.* Nature, 2016. **533**(7601): p. 58-63.
71. Eriksen, D.T., et al., *Directed evolution of a cellobiose utilization pathway in Saccharomyces cerevisiae by simultaneously engineering multiple proteins.* Microbial Cell Factories, 2013. **12**(1): p. 61.
72. Kim, J.H., et al., *High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice.* PLoS ONE, 2011. **6**(4): p. e18556.
73. Park, M., et al., *Expression of serotonin derivative synthetic genes on a single self-processing polypeptide and the production of serotonin derivatives in microbes.* Applied Microbiology and Biotechnology, 2008. **81**(1): p. 43-49.
74. Beekwilder, J., et al., *Polycistronic expression of a β-carotene biosynthetic pathway in Saccharomyces cerevisiae coupled to β-ionone production.* Journal of Biotechnology, (0).
75. Zhang, X., et al., *Metabolic evolution of energy-conserving pathways for succinate production in Escherichia coli.* Proceedings of the National Academy of Sciences, 2009. **106**(48): p. 20180-20185.
76. Unrean, P. and F. Srienc, *Metabolic networks evolve towards states of maximum entropy production.* Metabolic Engineering, 2011. **13**(6): p. 666-673.
77. Trinh, C.T. and F. Srienc, *Metabolic Engineering of Escherichia coli for Efficient Conversion of Glycerol to Ethanol.* Applied and Environmental Microbiology, 2009. **75**(21): p. 6696-6705.
78. Fong, S.S., et al., *In silico design and adaptive evolution of Escherichia coli for production of lactic acid.* Biotechnology and Bioengineering, 2005. **91**(5): p. 643-648.
79. Agresti, J.J., et al., *Ultrahigh-throughput screening in drop-based microfluidics for directed evolution.* Proceedings of the National Academy of Sciences, 2010.
80. Romero, P.A., T.M. Tran, and A.R. Abate, *Dissecting enzyme function with microfluidic-based deep mutational scanning.* Proceedings of the National Academy of Sciences, 2015. **112**(23): p. 7159-7164.

81. Chen, K. and F.H. Arnold, *Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(12): p. 5618-5622.
82. Stemmer, W.P.C., *Rapid evolution of a protein in vitro by DNA shuffling.* Nature, 1994. **370**(6488): p. 389-391.
83. Brustad, E.M., et al., *Structure-Guided Directed Evolution of Highly Selective P450-Based Magnetic Resonance Imaging Sensors for Dopamine and Serotonin.* Journal of Molecular Biology, 2012. **422**(2): p. 245-262.
84. Chen, Z. and H. Zhao, *Rapid Creation of a Novel Protein Function by in Vitro Coevolution.* Journal of Molecular Biology, 2005. **348**(5): p. 1273-1282.
85. Cho, G.S. and J.W. Szostak, *Directed Evolution of ATP Binding Proteins from a Zinc Finger Domain by Using mRNA Display.* Chemistry & Biology, 2006. **13**(2): p. 139-147.
86. Fasan, R., et al., *Evolutionary History of a Specialized P450 Propane Monooxygenase.* Journal of Molecular Biology, 2008. **383**(5): p. 1069-1080.
87. Wang, H.H., et al., *Programming cells by multiplex genome engineering and accelerated evolution.* Nature, 2009. **460**.
88. DiCarlo, J.E., et al., *Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.* Nucleic Acids Research, 2013.
89. Findlay, G.M., et al., *Saturation editing of genomic regions by multiplex homology-directed repair.* Nature, 2014. **513**(7516): p. 120-123.
90. Ravikumar, A., A. Arrieta, and C.C. Liu, *An orthogonal DNA replication system in yeast.* Nature Chemical Biology, 2014. **10**(3): p. 175-177.
91. Romanini, D.W., et al., *A Heritable Recombination System for Synthetic Darwinian Evolution in Yeast.* ACS Synthetic Biology, 2012. **1**(12): p. 602-609.
92. Wilhelm, F.X., M. Wilhelm, and A. Gabriel, *Reverse transcriptase and integrase of the Saccharomyces cerevisiae Ty1 element.* Cytogenetic and Genome Research, 2005. **110**(1-4): p. 269-287.
93. Boeke, J., H. Xu, and G. Fink, *A general method for the chromosomal amplification of genes in yeast.* Science, 1988. **239**(4837): p. 280-282.
94. Curcio, M.J. and D.J. Garfinkel, *Single-step selection for Ty1 element retrotransposition.* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(3): p. 936-940.
95. Gabriel, A., et al., *Replication infidelity during a single cycle of Ty1 retrotransposition.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(15): p. 7767-7771.
96. Boutabout, M., M. Wilhelm, and F.-X. Wilhelm, *DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposon Ty1.* Nucleic Acids Research, 2001. **29**(11): p. 2217-2222.
97. Holstege, F.C.P., et al., *Dissecting the Regulatory Circuitry of a Eukaryotic Genome.* Cell. **95**(5): p. 717-728.

98. Scholes, D.T., et al., *Multiple Regulators of Ty1 Transposition in Saccharomyces cerevisiae Have Conserved Roles in Genome Maintenance.* Genetics, 2001. **159**(4): p. 1449.
99. Chan, J.E. and R.D. Kolodner, *A Genetic and Structural Study of Genome Rearrangements Mediated by High Copy Repeat Ty1 Elements.* PLoS Genet, 2011. **7**(5): p. e1002089.
100. Qian, Z., et al., *Yeast Ty1 Retrotransposition Is Stimulated by a Synergistic Interaction between Mutations in Chromatin Assembly Factor I and Histone Regulatory Proteins.* Molecular and Cellular Biology, 1998. **18**(8): p. 4783-4792.
101. Luria, S.E. and M. Delbrück, *MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE.* Genetics, 1943. **28**(6): p. 491.
102. Paquin, C.E. and V.M. Williamson, *Temperature Effects on the Rate of Ty Transposition.* Science, 1984. **226**(4670): p. 53.
103. Curran, K.A., et al., *Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast.* ACS Synthetic Biology, 2015. **4**(7): p. 824-832.
104. Kim, C. and S. Mobashery, *Phosphoryl transfer by aminoglycoside 3′-phosphotransferases and manifestation of antibiotic resistance.* Bioorganic Chemistry, 2005. **33**(3): p. 149-158.
105. Esvelt, K.M., J.C. Carlson, and D.R. Liu, *A system for the continuous directed evolution of biomolecules.* Nature, 2011. **472**(7344): p. 499-U550.
106. Zhu, Y.O., et al., *Precise estimates of mutation rate and spectrum in yeast.* Proceedings of the National Academy of Sciences, USA, 2014. **111**(22): p. E2310-E2318.
107. Boeke, J.D., et al., *[10] 5-Fluoroorotic acid as a selective agent in yeast molecular genetics*, in *Methods in Enzymology*. 1987, Academic Press. p. 164-175.
108. Cormack, B.P. and K. Struhl, *The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells.* Cell, 1992. **69**(4): p. 685-696.
109. Young, E., S.-M. Lee, and H. Alper, *Optimizing pentose utilization in yeast: the need for novel tools and approaches.* Biotechnology for Biofuels, 2010. **3**(1): p. 24.
110. Träff, K.L., et al., *Deletion of the GRE3 Aldose Reductase Gene and Its Influence on Xylose Metabolism in Recombinant Strains of Saccharomyces cerevisiae Expressing the xylA and XKS1 Genes.* Applied and Environmental Microbiology, 2001. **67**(12): p. 5668-5674.
111. Blazeck, J., et al., *Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters.* Biotechnology and Bioengineering, 2012. **109**(11): p. 2884-2895.
112. Lee, W.K., *Cloned Gene Integration in Recombinant Yeast and Application to Metabolic Engineering*. 2003, University of California Irvine.
113. Nissley, D.V., D.J. Garfinkel, and J.N. Strathern, *HIV reverse transcription in yeast.* Nature, 1996. **380**(6569): p. 30-30.

114. Merkulov, G.V., et al., *Ty1 Proteolytic Cleavage Sites Are Required for Transposition: All Sites Are Not Created Equal.* Journal of Virology, 2001. **75**(2): p. 638-644.
115. Shah, F.S., et al., *Differential influence of nucleoside analog-resistance mutations K65R and L74V on the overall mutation rate and error specificity of human immunodeficiency virus type 1 reverse transcriptase.* Journal of Biological Chemistry, 2000.
116. Stumpp, S.N., B. Heyn, and S. Brakmann, *Activity-based selection of HIV-1 reverse transcriptase variants with decreased polymerization fidelity.* Biological Chemistry, 2010. **391.6**: p. 665-674.
117. Kaushik, N., et al., *Tyrosine 222, a Member of the YXDD Motif of MuLV RT, Is Catalytically Essential and Is a Major Component of the Fidelity Center†.* Biochemistry, 1999. **38**(9): p. 2617-2627.
118. Kaushik, N., et al., *Valine of the YVDD Motif of Moloney Murine Leukemia Virus Reverse Transcriptase: Role in the Fidelity of DNA Synthesis†.* Biochemistry, 2000. **39**(17): p. 5155-5165.
119. Halvas, E.K., E.S. Svarovskaia, and V.K. Pathak, *Role of Murine Leukemia Virus Reverse Transcriptase Deoxyribonucleoside Triphosphate-Binding Site in Retroviral Replication and In Vivo Fidelity.* Journal of Virology, 2000. **74**(22): p. 10349-10358.
120. Halvas, E.K., E.S. Svarovskaia, and V.K. Pathak, *Development of an In Vivo Assay To Identify Structural Determinants in Murine Leukemia Virus Reverse Transcriptase Important for Fidelity.* Journal of Virology, 2000. **74**(1): p. 312-319.
121. Crook, N., et al., *In vivo continuous evolution of genes and pathways in yeast.* Nature Communications, 2016. **7**: p. 13051.
122. Dapp, M.J., R.H. Heineman, and L.M. Mansky, *Interrelationship between HIV-1 Fitness and Mutation Rate.* Journal of Molecular Biology, 2013. **425**(1): p. 41-53.
123. Lwatula, C., S.J. Garforth, and V.R. Prasad, *Lys66 residue as a determinant of high mismatch extension and misinsertion rates of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT).* The FEBS journal, 2012. **279**(21): p. 4010-4024.
124. Curcio, M.J., S. Lutz, and P. Lesage, *The Ty1 LTR-Retrotransposon of Budding Yeast, Saccharomyces cerevisiae.* Microbiology Spectrum, 2015. **3**(2).
125. Youngren, S.D., et al., *Functional organization of the retrotransposon Ty from Saccharomyces cerevisiae: Ty protease is required for transposition.* Molecular and Cellular Biology, 1988. **8**(4): p. 1421-1431.
126. Wilhelm, M. and F.X. Wilhelm, *Role of Integrase in Reverse Transcription of the Saccharomyces cerevisiae Retrotransposon Ty1.* Eukaryotic Cell, 2005. **4**(6): p. 1057-1065.
127. BOLTON, E.C., et al., *Identification and characterization of critical cis-acting sequences within the yeast Ty1 retrotransposon.* RNA, 2005. **11**(3): p. 308-322.

128. Brakmann, S. and S. Grzeszik, *An Error-Prone T7 RNA Polymerase Mutant Generated by Directed Evolution.* ChemBioChem, 2001. **2**(3): p. 212-219.
129. Dower, K.E.N. and M. Rosbash, *T7 RNA polymerase-directed transcripts are processed in yeast and link 3′ end formation to mRNA nuclear export.* RNA, 2002. **8**(5): p. 686-697.
130. Curran, K.A., et al., *Use of expression-enhancing terminators in Saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications.* Metabolic Engineering, 2013. **19**(0): p. 88-97.
131. Mahr, R. and J. Frunzke, *Transcription factor-based biosensors in biotechnology: current state and future prospects.* Applied Microbiology and Biotechnology, 2016. **100**(1): p. 79-90.
132. Rogers, J.K., N.D. Taylor, and G.M. Church, *Biosensor-based engineering of biosynthetic pathways.* Current Opinion in Biotechnology, 2016. **42**: p. 84-91.
133. Paige, J.S., et al., *Fluorescence Imaging of Cellular Metabolites with RNA.* Science, 2012. **335**(6073): p. 1194.
134. McKeague, M. and M.C. DeRosa, *Challenges and Opportunities for Small Molecule Aptamer Development.* Journal of Nucleic Acids, 2012. **2012**: p. 20.
135. Lütke-Eversloh, T., C.N.S. Santos, and G. Stephanopoulos, *Perspectives of biotechnological production of l-tyrosine and its applications.* Applied Microbiology and Biotechnology, 2007. **77**(4): p. 751-762.
136. Mannironi, C., et al., *Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif.* RNA, 2000. **6**(4): p. 520-527.
137. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Research, 2003. **31**(13): p. 3406-3415.
138. Song, W., et al., *Plug-and-Play Fluorophores Extend the Spectral Properties of Spinach.* Journal of the American Chemical Society, 2014. **136**(4): p. 1198-1201.
139. Song, W., R.L. Strack, and S.R. Jaffrey, *Imaging bacterial protein expression using genetically encoded RNA sensors.* Nat Meth, 2013. **10**(9): p. 873-875.
140. Tawfik, D.S. and A.D. Griffiths, *Man-made cell-like compartments for molecular evolution.* Nat Biotech, 1998. **16**(7): p. 652-656.
141. Aharoni, A., et al., *High-Throughput Screening of Enzyme Libraries: Thiolactonases Evolved by Fluorescence-Activated Sorting of Single Cells in Emulsion Compartments.* Chemistry & Biology. **12**(12): p. 1281-1289.
142. Miller, O.J., et al., *Directed evolution by in vitro compartmentalization.* Nat Meth, 2006. **3**(7): p. 561-570.
143. Huebner, A., et al., *Quantitative detection of protein expression in single cells using droplet microfluidics.* Chemical Communications, 2007(12): p. 1218-1220.
144. Wang, B.L., et al., *Microfluidic high-throughput culturing of single cells for selection based on extracellular metabolite production or consumption.* Nat Biotech, 2014. **32**(5): p. 473-478.

145. Sciambi, A. and A.R. Abate, *Accurate microfluidic sorting of droplets at 30 kHz.* Lab on a Chip, 2015. **15**(1): p. 47-51.
146. Chen, B., et al., *High-throughput analysis and protein engineering using microcapillary arrays.* Nat Chem Biol, 2016. **12**(2): p. 76-81.
147. Hartmann, M., et al., *Evolution of feedback-inhibited β/α barrel isoenzymes by gene duplication and a single mutation.* Proceedings of the National Academy of Sciences, 2003. **100**(3): p. 862-867.
148. Luttik, M.A.H., et al., *Alleviation of feedback inhibition in Saccharomyces cerevisiae aromatic amino acid biosynthesis: Quantification of metabolic impact.* Metabolic Engineering, 2008. **10**(3–4): p. 141-153.
149. Lütke-Eversloh, T. and G. Stephanopoulos, *A semi-quantitative high-throughput screening method for microbial l-tyrosine production in microtiter plates.* Journal of Industrial Microbiology & Biotechnology, 2007. **34**(12): p. 807-811.
150. Hartmann, M., et al., *Evolution of feedback-inhibited β/α barrel isoenzymes by gene duplication and a single mutation.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(3): p. 862-867.
151. Wagner, J.M. and H.S. Alper, *Synthetic biology and molecular genetics in non-conventional yeasts: Current tools and future advances.* Fungal Genetics and Biology, 2016. **89**: p. 126-136.
152. crook, N., *Novel Approaches for Metabolic Engineering of Yeast at Multiple Scales*, in *Chemical Engineering*. 2014, University of Texas at Austin.
153. Arnold, C.E., et al., *Leader peptide efficiency correlates with signal recognition particle dependence in Saccharomyces cerevisiae.* Biotechnology and Bioengineering, 1998. **59**(3): p. 286-293.
154. Donnelly, M.L.L., et al., *The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences.* Journal of General Virology, 2001. **82**(5): p. 1027-1041.
155. Verwaal, R., et al., *High-Level Production of Beta-Carotene in Saccharomyces cerevisiae by Successive Transformation with Carotenogenic Genes from Xanthophyllomyces dendrorhous.* Applied and Environmental Microbiology, 2007. **73**(13): p. 4342-4350.
156. Li, H., O. Schmitz, and H.S. Alper, *Enabling glucose/xylose co-transport in yeast through the directed evolution of a sugar transporter.* Applied Microbiology and Biotechnology, 2016: p. 1-9.
157. Young, E., et al., *Functional Survey for Heterologous Sugar Transport Proteins, Using Saccharomyces cerevisiae as a Host.* Applied and Environmental Microbiology, 2011. **77**(10): p. 3311-3319.
158. Horwitz, Andrew A., et al., *Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas.* Cell Systems. **1**(1): p. 88-96.
159. Richard, P., K. Viljanen, and M. Penttilä, *Overexpression of PAD1 and FDC1 results in significant cinnamic acid decarboxylase activity in Saccharomyces cerevisiae.* AMB Express, 2015. **5**: p. 12.

160. Clarke, C.F. and C.M. Allan, *Biochemistry: Unexpected role for vitamin B2.* Nature, 2015. **522**(7557): p. 427-428.
161. Sambrook J, R.D., *Molecular Cloning: A Laboratory Manual, 3rd ed.* 2001: Cold Spring Harbor Laboratory Press.
162. Hegemann, J.H. and S.B. Heick, *Delete and Repeat: A Comprehensive Toolkit for Sequential Gene Knockout in the Budding Yeast Saccharomyces cerevisiae*, in *Strain Engineering: Methods and Protocols*, J.A. Williams, Editor. 2011, Humana Press: Totowa, NJ. p. 189-206.
163. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Meth, 2009. **6**(5): p. 343-345.
164. Lamprecht M, S.D., and Carpenter A, *CellProfiler™: free, versatile software for automated biological image analysis.* BioTechniques, 2007. **42**(1): p. 71-75.
165. Hall, B.M., et al., *Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria–Delbrück fluctuation analysis.* Bioinformatics, 2009. **25**(12): p. 1564-1565.
166. Ma, W.T., G.V. Sandri, and S. Sarkar, *Analysis of the Luria-Delbrück Distribution Using Discrete Convolution Powers.* Journal of Applied Probability, 1992. **29**(2): p. 255-267.
167. Masella, A.P., et al., *PANDAseq: paired-end assembler for illumina sequences.* BMC Bioinformatics, 2012. **13**(1): p. 31.
168. najoshi. *https://github.com/najoshi/sabre.git*. Available from: https://github.com/najoshi/sabre.git.
169. Patel, R.K. and M. Jain, *NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data.* PLoS ONE, 2012. **7**(2): p. e30619.
170. Ning, Z., A.J. Cox, and J.C. Mullikin, *SSAHA: A Fast Search Method for Large DNA Databases.* Genome Research, 2001. **11**(10): p. 1725-1729.
171. CLOPPER, C.J. and E.S. PEARSON, *THE USE OF CONFIDENCE OR FIDUCIAL LIMITS ILLUSTRATED IN THE CASE OF THE BINOMIAL.* Biometrika, 1934. **26**(4): p. 404-413.
172. Zaccolo, M., et al., *An Approach to Random Mutagenesis of DNA Using Mixtures of Triphosphate Derivatives of Nucleoside Analogues.* Journal of Molecular Biology, 1996. **255**(4): p. 589-603.